Paleoceanography and Paleoclimatology

RESEARCH ARTICLE
10.1029/2024PA004991

Statistical Fingerprints of Forced and Unforced Variability Reveal Inconsistencies Between Marine Proxies and Climate Models on Multi-Decadal to Millennial Timescales

Rebecca Cleveland Stout[1], **Cristian Proistosescu**[2], **and Gerard Roe**[1,3]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, [2]Department of Climate, Meteorology & Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana-Champaign, IL, USA, [3]Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA

Key Points:

- Forced and unforced models have distinct statistical fingerprints of temperature variability, defined using spectral slope and coherence
- Holocene marine sediment cores have steep spectral slopes and low spatial correlations on timescales >400 years, contradicting model behavior
- Our results identify targets to ensure the dynamic consistency of proxy records with one another and with theoretical expectations

Supporting Information:

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

Correspondence to:

R. Cleveland Stout,
rrcs@uw.edu

Citation:

Cleveland Stout, R., Proistosescu, C., & Roe, G. (2025). Statistical fingerprints of forced and unforced variability reveal inconsistencies between marine proxies and climate models on multi-decadal to millennial timescales. *Paleoceanography and Paleoclimatology*, *40*, e2024PA004991. https://doi.org/10.1029/2024PA004991

Received 20 AUG 2024
Accepted 27 AUG 2025

Author Contributions:

Conceptualization: Rebecca Cleveland Stout, Cristian Proistosescu, Gerard Roe
Data curation: Rebecca Cleveland Stout
Formal analysis: Rebecca Cleveland Stout
Funding acquisition: Rebecca Cleveland Stout, Cristian Proistosescu, Gerard Roe
Investigation: Rebecca Cleveland Stout, Cristian Proistosescu, Gerard Roe
Methodology: Rebecca Cleveland Stout, Cristian Proistosescu, Gerard Roe
Project administration: Cristian Proistosescu, Gerard Roe
Supervision: Cristian Proistosescu, Gerard Roe
Validation: Rebecca Cleveland Stout
Visualization: Rebecca Cleveland Stout
Writing – original draft: Rebecca Cleveland Stout

Abstract Comparing general circulation models (GCMs) and Holocene paleoclimate proxies reveals large mismatches between simulated and reconstructed low-frequency temperature variability on multidecadal to multicentennial timescales. These mismatches—stemming from proxy-related artifacts or model deficiencies in simulating forced or unforced variability—impact future warming predictions and past climate interpretations. Theory and models suggest forced and unforced variability can be statistically fingerprinted based on robust differences in their spatial pattern and magnitude on these long timescales. This motivates us to characterize frequency-dependent behavior across 52 globally-distributed Holocene sediment cores with Mg/Ca- and $U^{K'}_{37}$-based temperature reconstructions and a suite of forced and unforced GCM simulations using two metrics: spectral slope β and cross-coherence. Compared with unforced simulations, forced simulations exhibit steeper β and higher cross-coherence on centennial-and-longer timescales, especially in regions with applied forcing. Proxy-system models suggest Mg/Ca and $U^{K'}_{37}$-based temperature reconstructions can preserve some statistical differences between forced and unforced β and coherence. We find inconsistencies between models and proxies, with proxies having higher β, indicating more variability (forced or unforced), but lower coherence between sites, indicating more unforced variability. Our analyses highlight: (a) coherence among nearby sites as an important measure of the regional and large-scale information retained in those records, (b) the need for a larger set of long climate simulations to probe the range of model behavior on these timescales, and (c) the need to integrate proxy-system models with GCM output to evaluate performance of both proxy networks and climate models.

Plain Language Summary Global-mean temperature can change due to factors external to the climate system (e.g., greenhouse-gases) or due to factors internal to the climate system (e.g., energy moving between the ocean and atmosphere). While observations constrain short-term variability, understanding long-term changes requires paleoclimate records (proxy reconstructions) and models. On decadal-plus timescales, proxy reconstructions show more variability than climate models simulate. Whether this extra variability is externally forced or unforced remains an open question. Models suggest forced and unforced variability have distinct "statistical fingerprints," and differ in magnitude, spatial pattern, and coherence–in other words, the strength of temperature fluctuations and the degree of shared variability across locations depend on both geographic position and the presence of external forcing. Using two frequency-dependent metrics, we assess whether marine sediment proxies align with forced or unforced variability. We find inconsistencies, as marine proxies exhibit characteristics of both. Constraining the statistical fingerprints of proxy temperature variability, including spatial patterns, is crucial for understanding Holocene temperature drivers.

1. Introduction

Understanding the basic behavior of the climate system rests upon our ability to constrain the spatial pattern of temperature variability across timescales. Changes in temperature can be either forced or unforced. Forced changes result from external radiative forcings imposed at the top of the atmosphere (e.g., volcanic aerosols, land-use changes, changes in greenhouse-gas concentrations, or changes in incoming solar radiation; Bindoff et al., 2013). Unforced changes result from the internal dynamics of the climate system (Hasselmann, 1976; Hawkins & Sutton, 2009; Leith, 1978), involving a redistribution of heat between different reservoirs (e.g., heat moving between the ocean and atmosphere; Meehl et al., 2013) or internal changes in the total heat in the climate system (e.g., internal changes in clouds or sea ice; Brown et al., 2014; Palmer & McNeall, 2014). Variations in

Writing – review & editing: Rebecca Cleveland Stout, Cristian Proistosescu, Gerard Roe

temperature can thus be partitioned into a "forced" (external) component and an "unforced" (internal) component, although there may also be nonlinear interactions between the two (e.g., O'Brien & Deser, 2023). Constraining variability across timescales can help illuminate drivers of variability—that is, determine how much of that variability is forced or unforced—at a given timescale, and underpins predictions of future temperature change under anthropogenic warming (e.g., Laepple et al., 2023; Solomon et al., 2011; von der Heydt et al., 2021).

Our ability to accurately simulate temperature variability depends on the timescale of interest. On seasonal-to-interannual timescales, variability produced by coupled general-circulation models (GCMs) is generally consistent with observational data on a global scale (Hausfather et al., 2020; Jones et al., 2013), although some regional model-data discrepancies have been noted (M. Collins et al., 2001; Min et al., 2005). The mechanisms of global-mean temperature variability on these timescales are also relatively well-understood, with the El-Niño Southern Oscillation playing a dominant role (e.g., Brown et al., 2015; Foster & Rahmstorf, 2011). However, on interdecadal-and-longer timescales, different GCMs have different magnitudes of global-mean temperature variability, with different regions responsible for driving variability (Brown et al., 2015; Parsons & Hakim, 2019; Parsons et al., 2020; Wills et al., 2021). The instrumental record is too short ($\leq$150 years of data in most locations) to constrain this variability (Neukom et al., 2019); thus, GCM estimates of climate variability must be tested against our best reconstructions of past climate.

Of particular interest here are the slowest modes of adjustment, which act on multicentennial-to-millennial timescales. These slow modes will play an increasingly important role in future temperature evolution. For example, Proistosescu and Huybers (2017) showed that approximately 40% of the response to anthropogenic forcing occurs on a dynamical mode with timescales >300 years. Consistent with this, Rugenstein et al. (2020) showed a continued evolution of global-mean temperature in coupled climate models even many centuries after a step-change doubling of atmospheric CO_2. Constraining these slow modes from natural variability requires long data sets, ideally many times longer than the timescale of adjustment.

Holocene climate has the best potential to constrain low-frequency climate variability because: (a) proxy data from the last few millennium has relatively high resolution in time and space (e.g., PAGES2k Consortium, 2017); and (b) the climate system (including mean temperature and boundary conditions) over this period was similar to present, and thus can provide meaningful insight into present-day climate variability. There has recently been an explosion of work on statistical characterizations of climate variability over the recent Holocene. These efforts include curating and aggregating proxy data (PAGES2k Consortium, 2017); developing proxy-system models, which comprise of statistical or process-based equations that predict proxy responses from GCM output and allow direct comparisons between proxy records and model simulations (e.g., Dee et al., 2017); and creating fused model-data products such as the Last Millennium Reanalysis (Hakim et al., 2016; Tardif et al., 2019).

Composite Holocene proxy records, GCMs, and data-assimilation products largely agree on the magnitude of global temperature variability (Laepple et al., 2023; Zhu et al., 2019). However, GCMs appear to underestimate past low-frequency climate variability relative to Holocene proxies on local and regional spatial scales (Ault, Cole, et al., 2013; Laepple & Huybers, 2014b; Laepple et al., 2023). The source of unexplained low-frequency variability in proxy records is an open question, with critical implications for both interpreting past climate variability and predicting future climate change. Here, we will explore three possible explanations: (a) GCMs underestimate forced variability, either by underestimating forcings or underestimating the climate's response to those forcing; (b) GCMs have insufficient internal variability at low frequencies; or (c) proxy records contain biases and noise that amplify measured low-frequency variability. In this study, we use "forced variability" to include temperature changes resulting from long-term boundary condition variations, including ice-sheet changes.

In principle, the "statistical fingerprint" of temperature variability—defined by its magnitude and spatial pattern and characterized using frequency-dependent metrics—at low frequencies is strongly related to whether the variability is forced or unforced. This distinction can aid in interpreting past temperature records. Based on insights from GCMs and idealized models, when variability is forced, the magnitude of variability is higher, and the number of independent spatial patterns of temperature is significantly reduced (Cleveland Stout et al., 2023). We are interested in whether proxy records are dynamically consistent with model findings, and if low-frequency variability in the proxy record can primarily be characterized as forced or unforced. To that end, we characterize forced and unforced variability across a suite of GCMs using two frequency-dependent metrics: (a) spectral slope ($\Subset \beta$), which characterizes the relative magnitude of low-frequency variability and high-frequency

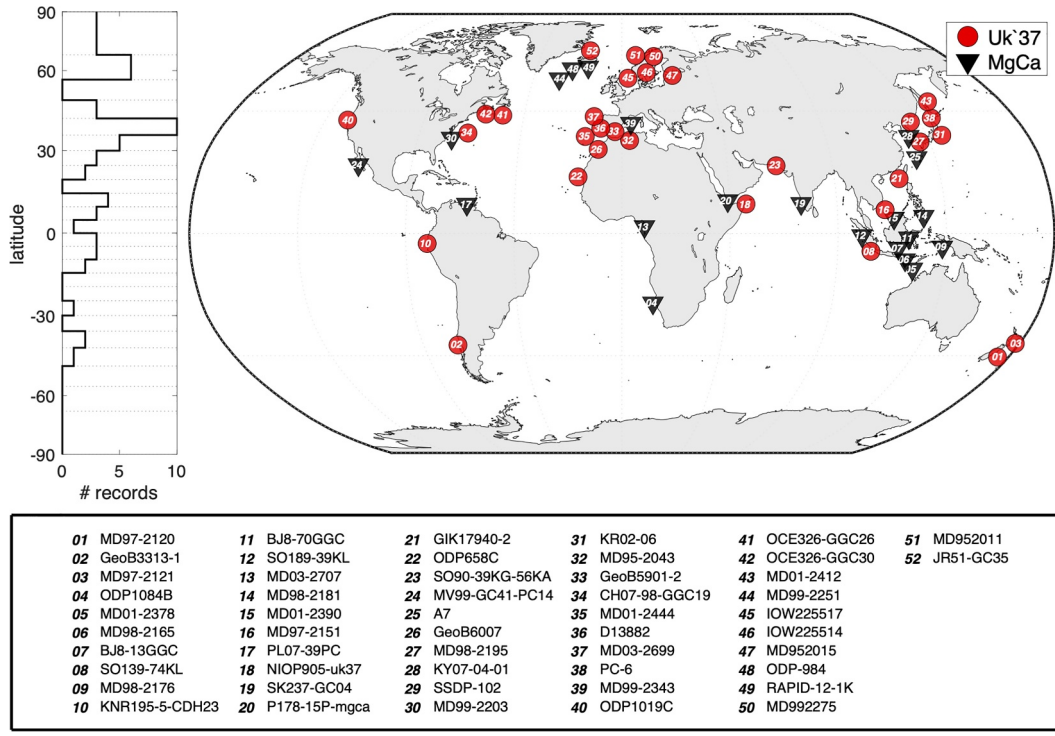

Figure 1. Locations of marine sediment cores used in this study. Colors and shapes indicate the type of proxy used to reconstruct local Holocene sea-surface temperature; red circles are alkenone-based reconstructions and black triangles are MgCa-based reconstructions. Where records overlap, markers are slightly geographically repositioned for readability. Markers are numbered according to latitude (1, furthest South; 52, furthest North). On the left, the geographic distribution of proxy records is shown (number of total proxy records in each of 24 equal-area bins).

variability; and (b) coherence, which measures the shared variability between temperature at different locations. These metrics (which comprise a "statistical fingerprint") constrain the magnitude of variability (β) and the spatial pattern of variability (coherence and the meridional structure of β). We then perform similar analyses on a set of marine sediment cores with relatively high sedimentation rates, and which contain climate information from the past ~10,000 years based on two proxy temperature archives, Mg/Ca ratios and the alkenone unsaturation index $U^{K'}_{37}$. Finally, we perform a pseudoproxy experiment–creating hypothetical proxy records based on modeled temperature–to explore the degree to which proxy noise and bias may have corrupted the measured signals of coherence between different locations. Past work characterized a proxy-model mismatch in low-frequency variability that was especially pronounced in the tropics (30°S–30°N) (Laepple & Huybers, 2014b). Here, we build on that work by exploring how the absence or presence of forcing influences the spatial pattern of low-frequency variability, which may offer new insights into the mechanisms driving proxy-model discrepancies in variability. We also perform our analysis on an expanded proxy database. Ultimately, we are motivated to characterize the forced and unforced components of slow modes of climate adjustment across the Holocene to empirically constrain multidecadal-and-longer timescale variability.

2. Data and Methods

2.1. Proxy Temperature Reconstructions

We build a database of Holocene sediment cores with relatively high sedimentation rates. This database combines two existing compilations from Laepple and Huybers (2014a) and Rehfeld et al. (2018), and also adds recently-published data sets from four sites: MD99-2343, MD95-2043 (Català et al., 2019), MD95-2015, and MD99-2275 (Sicre et al., 2021). We filter the database for records with mean temporal resolution of less than 200 years, which yields 52 records from −45.5°S to 67°N (Figure 1). These records are primarily located in the Northern Hemisphere (40/52 are north of the equator), and most come from either continental margins—which may exhibit offsets from zonal-mean sea-surface temperature (SST) values due to gyre circulations—or epeiric seas (e.g., the

Mediterranean and North Seas)—which tend to be warmer and have higher seasonality than the open ocean (Judd et al., 2020). The mean temporal resolution $\left(\equiv \overline{\Delta t}\right)$ varies from $\overline{\Delta t}$ = 4 years (MD99-2275) to $\overline{\Delta t}$ = 198 years (PL07-39PC). This resolution allows us to estimate spectral behavior on timescales as short as 8–400 years, depending on the site. We truncate records to the past 10.5 ka calendar years before present (ka cal BP; BP is defined as prior to 1950). The youngest records extend to ∼2000 CE, with a mean record length of ∼7,080 years. Most ages are calculated using radiocarbon (^{14}C) dating, although there are a few exceptions where age models are built using correlations with composite δ^{18}O records or stratigraphic events with detailed radiocarbon chronologies. Additional details on the calibration and dating techniques can be found in the original sources (Català et al., 2019; Kim et al., 2004; Laepple & Huybers, 2013; Leduc et al., 2010a; Rehfeld et al., 2018; Sicre et al., 2021).

Of the total 52 records, 31 are alkenone-based reconstructions (Prahl & Wakeham, 1987), and 21 are Mg/Ca-based (Nürnberg et al., 1996). All records are converted from raw proxy units to SST (°C) using a standardized, global calibration to facilitate spatial comparisons. $U^{K'}_{37}$ records are recalibrated using 0.033 $U^{K'}_{37}$ per °C, U^{K}_{37} records are recalibrated using 0.035 U^{K}_{37} per °C, and Mg/Ca records are recalibrated using 9% Mg/Ca change per °C, following Müller et al. (1998) and Anand et al. (2003) for $U^{K'}_{37}$ and Mg/Ca, respectively. More recent global calibrations for both proxies exist; for Mg/Ca, calibration adjustments include accounting for salinity sensitivity, species-specific seasonal production and effects of post-depositional calcite dissolution (Tierney et al., 2019). For $U^{K'}_{37}$, new calibrations account for a saturation effect with diminished alkenone temperature sensitivity above ∼23°C, which would impact seven of our 52 sites (Tierney & Tingley, 2018). However, sensitivity tests show that different calibrations do not significantly influence estimated frequency-dependent behavior (Figure S1 in Supporting Information S1).

There are a few notable features of the SST reconstructions (Figure 2). Firstly, the Mg/Ca-based SST reconstructions exhibit much higher variance than the $U^{K'}_{37}$-based SST reconstructions, which Laepple and Huybers (2013) primarily attributed to the aliasing of seasonal and interannual variability, although other factors like calibration effects, oceanographic settings, and organism habitat may also influence proxy variability (e.g., Conte et al., 2006; Tierney & Tingley, 2018; Tierney et al., 2019). Secondly, the $U^{K'}_{37}$-based proxies show strong Holocene trends in SST. With some site-specific exceptions, over the past 10 kyr, sites at the low-latitudes generally indicate warming (sites 8, 10, 18, 21, 23–25) and sites at the high-latitudes—particularly in the North Atlantic—generally indicate cooling (sites 41–48; also see Figure S3 in Supporting Information S1). These North Atlantic trends are generally in agreement with temperature-sensitive records in Greenland, which show cooling of ∼3°C over the Holocene (Briner et al., 2016). Leduc et al. (2010a) argued that these trends may be an artifact of seasonal changes driven by precession, although systematic calibration biases at warm temperatures could also influence these trends (e.g., Conte et al., 2006; Tierney & Tingley, 2018). Mg/Ca-based reconstructions exhibit weaker SST trends since 10.5 ka cal BP than the $U^{K'}_{37}$-based reconstructions (Figure 2, black triangles cf. red circles); however, Mg/Ca paleothermometry relies on measuring Mg/Ca in a small number of foraminifera samples, which can exhibit large sample-to-sample spread, hindering the interpretation of these weaker trends (Leduc et al., 2010a).

2.2. Model Simulations

We are interested in understanding how internal and external variability might be represented in the proxy record. To identify features of internal and external variability, we rely on two experiments from CMIP5: the preindustrial control experiment (piControl), and the last-millennium experiment (past1000). Each of these experiments contains simulations from nine models. The piControl experiment simulates the unforced climate state (all external boundary conditions—greenhouse gas concentrations, aerosols, ozone, and solar irradiance—are held constant at preindustrial levels) for at least 400 years, and the past1000 experiment simulates the climate system from 850 to 1850 using transient external forcings (e.g., volcanic eruptions, vegetation changes, evolving greenhouse-gas concentrations). These are the same two experiments used in a previous study (Cleveland Stout et al., 2023). Monthly mean surface temperature output is annually averaged and regridded to a 128 × 64 grid (approximately 2.8° resolution). We note that the piControl experiment uses preindustrial boundary conditions, and boundary conditions may influence the spectral characteristics of unforced variability. Reschke et al. (2021), for example, found that models simulate higher spatial correlations during the Last Glacial Maximum than during

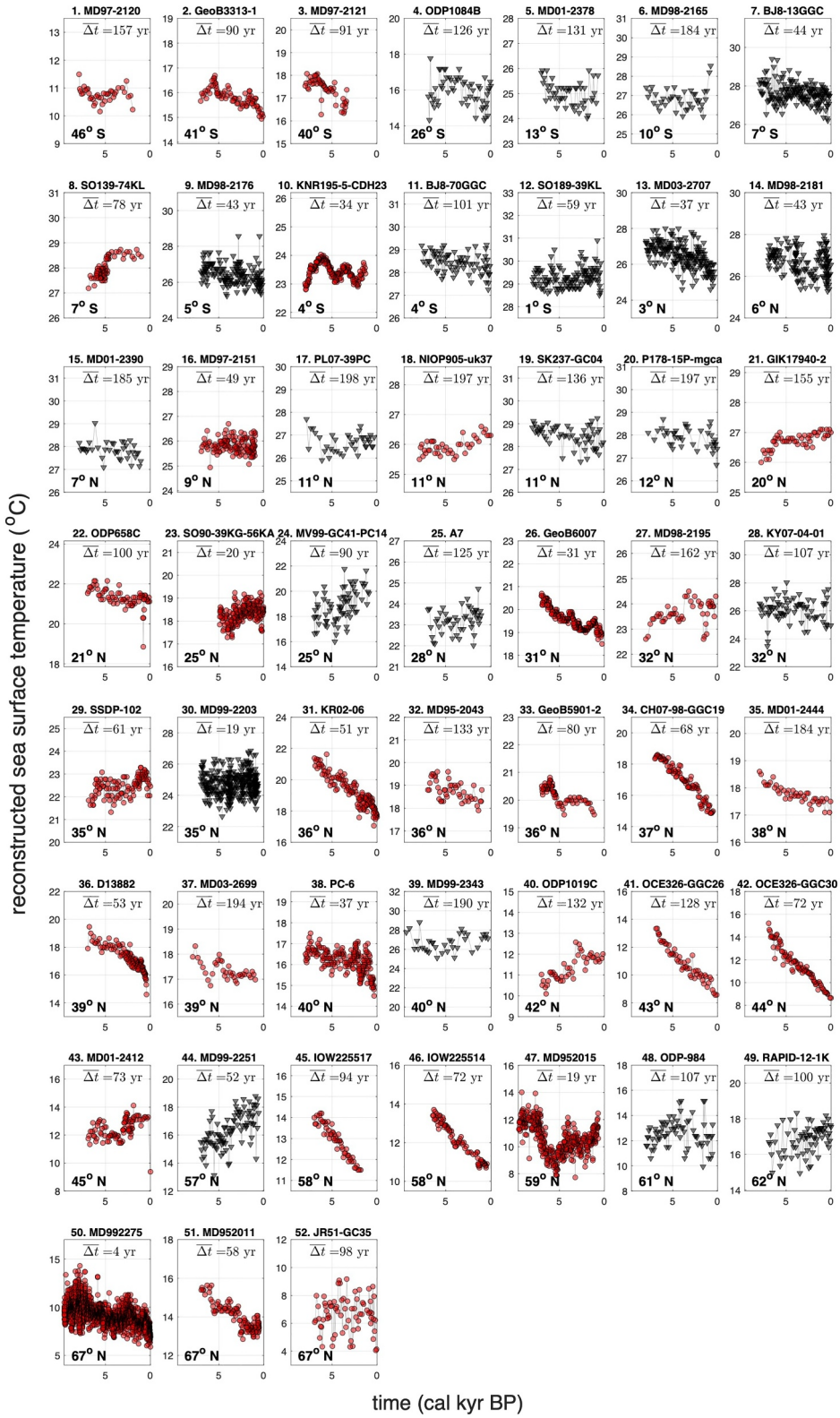

Figure 2. Sea-surface temperature reconstructions from proxy records. Colors and shapes indicate the type of proxy (red circles are $U^{K'}_{37}$-based reconstructions and black triangles are Mg/Ca-based reconstructions). Numbering is consistent with Figure 1. Latitude of record (rounded to the nearest degree) is shown in the bottom-left corner, and mean temporal resolution is shown in the top-right corner.

the Holocene, and Rehfeld et al. (2018) used proxy evidence to show that temperature variability was higher during the Last Glacial Maximum than during the Holocene. We do not consider this possible state-dependence. However, because much of the period of interest (10.5 ka to 2000 CE) had relatively stable global mean temperatures (e.g., Kaufman et al., 2020), we assume that approximating unforced variability under piControl boundary conditions is representative for the entire period.

The relatively short length of the CMIP5 simulations limits the direct comparison with spectral behavior in the proxy record. The highest resolvable frequencies in the proxy record are ∼1/4–1/400 years, depending on the site, and the lowest resolvable frequencies in the CMIP5 simulations are ∼1/400–1/1,200 years, depending on the model and experiment. We thus are primarily interested in using the piControl and past1000 experiments to understand how external forcing influences spectral behavior, and must use another tool to estimate modeled variability at millennial timescales.

We use output from TraCE-21ka to estimate modeled variability on these longer timescales that the proxy record primarily resolves. TraCE-21ka is a forced, transient climate simulation of the last 21,000 years using CCSM3 (He, 2011). Monthly mean surface temperature output is annually averaged and kept on its native grid (96 × 48, corresponding to 3.75° resolution). TraCE-21ka includes long-term forcings and evolving boundary conditions: variations in solar radiation driven by orbital forcing; changes in atmospheric greenhouse-gases; evolving ice sheets and paleogeography; and meltwater fluxes from retreating ice sheets. TraCE also has single-forcing experiments available, four of which we use: ORB, which includes transient orbital forcing; GHG, which includes transient greenhouse gas forcing; MWF, which includes Northern Hemisphere meltwater fluxes; and ICE, which includes evolving continental ice sheets. All other forcings are held constant. Results from the single-forcing experiments are shown in the supplement (Figures S7 and S8 in Supporting Information S1). We only use TraCE output from 10.5 ka cal BP to present to overlap with our proxy records. Over this period in the model, there is a 500-year-long meltwater flux imposed in the Northern Hemisphere (9.7–9.2 ka cal BP; flux is imposed in the Arctic, St. Lawrence River, and Hudson Strait), and a 5,500-year-long meltwater flux imposed in the Southern Hemisphere (10.5–5 ka cal BP; flux is imposed in the Ross and Weddell Seas). Importantly, TraCE-21ka does not include volcanic eruptions, which are the primary drivers of low-frequency variability in the past1000 simulations, nor does it include vegetation changes, which can produce significant regional and global-mean temperature fluctuations over the Holocene (Thompson et al., 2022). While other Holocene transient simulations are available, TraCE-21ka is used because it has more low-frequency temperature variability than any other model except EC-Earth3-LR (Askjær et al., 2022; Zhang et al., 2021), and thus offers an "upper bound" on the magnitude of modeled temperature variability.

We note that TraCE-21ka deviates from proxy SST reconstructions when comparing (a) mean temperature, and (b) linear trends since 10.5 ka cal BP (Figure S3 in Supporting Information S1). In particular, there are large cold biases in TraCE-21ka—particularly in the North Atlantic—which have been previously documented in a comparison with Greenland ice cores (Karger et al., 2023). TraCE-21ka is also colder than other Holocene transient simulations (e.g., ∼1°C colder than HadCM3 + SolVolcKK10) (Askjær et al., 2022). It is unclear why TraCE-21ka is colder than the SST reconstructions and why TraCE-21ka displays opposite temperature trends to proxies at some sites, although we note that CCSM3 has substantial cold biases (on the order of 1°C) in its representation of present-day North Atlantic SST as well (W. D. Collins et al., 2006). Many models struggle to produce unbiased estimates of many climatic processes at local and regional scales (e.g., M. Collins et al., 2013), and it is also possible that there are mean-state biases in the interpretation of Mg/Ca and $U_{37}^{K'}$ proxies that exacerbate model-data differences in SST (e.g., the error of the intercept term in a linear calibration) (Dolman et al., 2021). In addition, CCSM3 has consistent SST biases in coastal and upwelling regions (Large & Danabasoglu, 2006) which produces more homogenous temperature changes in response to Last Glacial Maximum forcing, relative to proxies, and thus could underestimate regional temperature variability (B. L. Otto-Bliesner et al., 2009). These biases persist even in later generations of the model (CESM) (Ma et al., 2019), and may be related to model resolution (Xu et al., 2022). Because most of the proxies are located on continental margins—where SST biases are particularly pronounced—comparisons between proxy variability and model variability at the grid-cell level must be interpreted with care. Despite these biases, global-mean temperature simulated in TraCE-21ka displays similar spectral behavior (power spectrum and spectral slope; see Section 2.3) to global-mean temperature estimates from observations, proxy records, and data assimilation products (Zhu et al., 2019), which motivates our use of TraCE-21ka.

We interpolate each model in the CMIP5 experiments (piControl and past1000) to a common 128 × 64 grid (approximately 2.8° resolution). Output from TraCE-21ka is kept on its native resolution (48 × 96 grid, approximately 3.7° resolution). When comparing between proxy records and simulations, for all models, we only use model output at the grid cells where proxy records are located. Because of the proximity between some proxy sites and the relatively coarse resolution of the model, the 52 proxy records are located in 48 model grid cells in the CMIP5 experiments, and 47 model grid cells in TraCE-21ka.

2.3. Estimation of Frequency-Dependent Behavior

There is evidence from observational data and proxy records that temperature can be characterized with a power-law spectrum (e.g., Hurst, 1951; Huybers & Curry, 2006; Pelletier & Turcotte, 1997),

$$S(f) \sim f^{-\beta} \tag{1}$$

Here, f is the frequency, the slope β is the strength of the scaling and $S(f)$ is the power spectral density. Power-law relationships arise due to nonlinearities in the physical climate system, although linear processes can also display power-law behavior over some frequencies (e.g., the power spectrum of an first-order autoregressive process has a non-zero slope at high frequencies, which flattens on low frequencies). A high, positive value of β indicates a "red" spectrum, with more variance on long timescales; a negative value of β implies a "blue" spectrum with more variance on short timescales; and $\beta = 0$ implies a "white" spectrum, with uniform variance across timescales (e.g., Ault, Deser, et al., 2013; Dee et al., 2017; Franzke et al., 2020; Nilsen et al., 2016; Zhu et al., 2019). Estimating β using local temperature timeseries allows us to evaluate the memory in the climate system at each site, with higher β indicating more memory (Lerch, 2012). We calculate spectral slopes based on proxy reconstructions (β_{proxy}), unforced simulations $(\beta_{\text{piControl}})$, and our two sets of forced simulations (past1000; β_{past1000}, and TraCE-21ka; β_{TraCE}).

Estimating β necessitates estimating the power spectrum $S(f)$ of our climate variable (in this case, temperature) at a given location. Most spectral estimators assume that data is regularly sampled in time, which does not hold true for most paleoclimate data. Traditionally, either the data is interpolated to a regular timestep prior to spectral analysis, or spectral analysis is performed on the original, temporally-uneven data using a Lomb-Scargle periodogram (equivalent to least-squares fitting of sine waves). However, both of these approaches can introduce bias: interpolation can bias spectral estimation (Rhines & Huybers, 2011; Wilson et al., 2003); and the periodogram is biased for finite samples and can exhibit spectral leakage. We instead use the multitaper-Lomb-Scargle (Springford et al., 2020), an approach that applies the Lomb-Scargle technique to the Thompson multitaper, which is less biased than the periodogram (Thomson, 1982). Estimated power spectra of proxy records are shown in Figure S2 in Supporting Information S1. To estimate power spectra of simulated temperature (for piControl, past1000, and TraCE-21ka experiments), we use the Thompson multi-taper (Thomson, 1982) because the output is evenly sampled in time. Timeseries are linearly detrended prior to all analysis to eliminate spurious power at the lowest frequencies. Detrending also allows us to ignore the apparent mismatch in Holocene temperature trends between proxies and models, which display long-term cooling and warming, respectively (e.g., Liu et al., 2014). In sensitivity tests with model timeseries, higher-order (up to fourth-order polynomial) detrending does not strongly influence our results. For all spectral estimations, three windows are used. β is calculated as the least-squares regression of the log-transformed power spectral density and log-transformed frequency. Following Huybers and Curry (2006), Dee et al. (2017), Ault, Cole, et al. (2013), we bin spectral densities (logarithmically spaced; 5 bins per log space) to avoid overweighting of high-frequency variability in calculating β. For the i^{th} record, the Nyquist frequency N_i (maximum cutoff frequency, or highest frequency that can be estimated) is calculated using the mean sampling frequency $\overline{\Delta f_{s_i}}$ of that record:

$$N_i = \frac{1}{2}\overline{\Delta f_{s_i}} = \frac{1}{2\overline{\Delta t_i}} \tag{2}$$

We do not show spectral estimates for the lowest 3 frequencies because the multitaper can be biased at low frequencies due to finite sample length (Laepple & Huybers, 2014b).

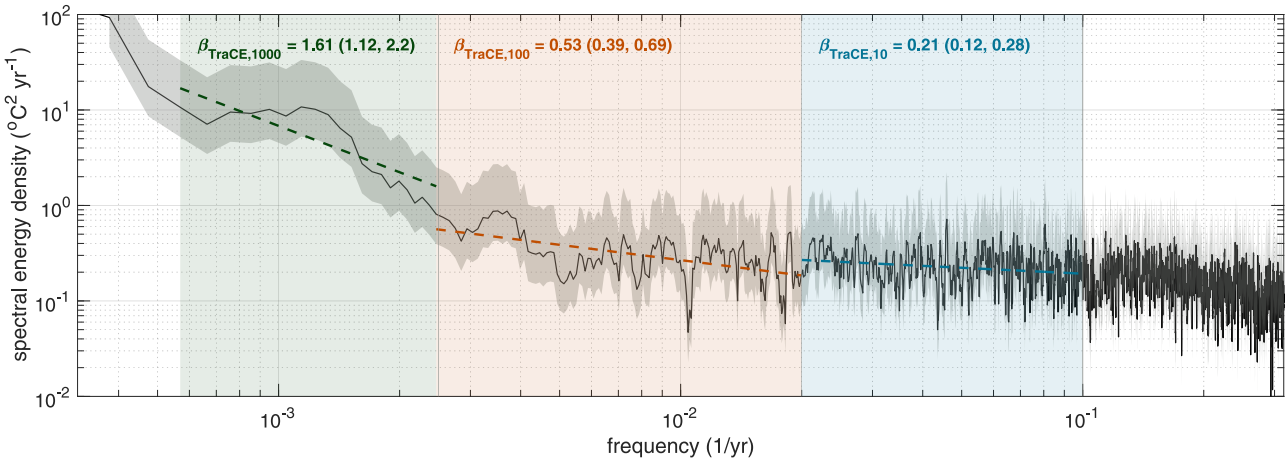

Figure 3. Spectral slope (β) estimation. Power spectrum of temperature at site 26 (GeoB6007; located at 30.85°N and 10.27°W) simulated in TraCE-21ka is shown in black. Spectral slopes are calculated in three frequency bands: 10–50 years (β_{10}, shaded blue), 50–400 years (β_{100}, shaded orange), and 400–2,000 years (β_{1000}, shaded green). Best-fit lines are shown in dashed colors, with the estimated β (with 95% uncertainty) shown in the top-left corner of each frequency band. Gray shading indicates the power spectrum 95% confidence interval, estimated with a chi-squared distribution.

There is evidence of multiple scaling regimes (frequency bands of at least an order of magnitude for which β is approximately constant) in regional temperature records and model simulations (e.g., Huybers & Curry, 2006; Zhu et al., 2019), although some work argues for a single scaling regime over the Holocene (Nilsen et al., 2016). The transitions between regimes vary depending on the model and observational record. Here, we choose to estimate β for three frequency bands: a multidecadal window (10–50 years; β_{10}), a multidecadal-to-multicentennial window (50–400 years; β_{100}); and a multicentennial-to-millennial window (400–2,000 years; β_{1000}). Estimating β in different frequency bands accounts for potential breaks in the spectral slope, which could arise from nonlinearities transferring energy between different modes of variability (Huybers & Curry, 2006), or a separation of mixing timescales for different energy reservoirs in the climate system, for example, between the mixed layer and the deep ocean (e.g., Fraedrich et al., 2004). An illustrative β calculation is shown in Figure 3. The frequency bins are chosen to be broadly consistent with the breaks in β seen in TraCE-21ka, although our results are not strongly dependent on the cutoff frequencies for each bin. We note that by calculating β over multiple scaling regimes, a temperature timeseries need not follow a power-law scaling over all frequencies; for example, a first-order autoregressive process can also be characterized with multiple β values.

We are also interested in evaluating similarities between temperature at different sites. One way of measuring this is with cross-coherence, which describes the correlation between temperature at two sites at a given frequency, and is measured between 0 (no correlation) and 1 (perfect direct relationship). Previous work shows that external forcing adds globally coherent signals of variability, which increase cross-coherence among sites, particularly on long timescales. When no external forcing is present, specific modes of internal variability—like the El-Niño Southern Oscillation—may produce robust global coherence among sites in a small frequency window (e.g., Mann & Park, 1993); however, the background climate may not be globally coherent even on multicentennial timescales, although some regional coherence may exist (Cleveland Stout et al., 2023). We characterize cross-coherence between temperature at each site, for both the proxy records and all model simulations. As for the power-spectra estimation, all timeseries are linearly detrended prior to analysis. As far as we are aware, no method yet exists for applying the Lomb-Scargle method to compute coherence. Because different records have different time resolutions, to estimate cross-coherence between sites i and j, we first interpolate the data to annual resolution, and then use a Thompson multi-taper to estimate coherence following (Mann & Park, 1993). We use three windows and a maximum frequency cutoff of $\min[N_i, N_j]$, discarding all coherence estimates at higher frequencies. Again, the coherence estimates at the lowest three frequencies are discarded to avoid bias. In experiments with synthetic timeseries, coherence biases from interpolation tend to be largest at high frequencies (close to the Nyquist frequency) and approach zero on low frequencies. We note that although (a) interpolating the original proxy data may bias coherence estimates and (b) some biases may also still exist at high frequencies for the Lomb-Scargle multitaper which could impact $\beta_{10,100}$, we use the same approaches to estimate coherence and spectral slope in the pseudo-proxies, so our pseudo-proxies would be biased in the same way.

2.4. Pseudo-Proxy Network

Proxy archives are imperfect recorders of past climate; temperature signals may be filtered and distorted by the proxy, making direct comparisons between proxy-based temperature reconstructions and model output difficult (e.g., Dee et al., 2017). Such comparisons can be made using proxy-system models, which simulate the proxy response to environmental forcing, and can be run with climate model inputs to produce "pseudo-proxies" (Dee et al., 2015; Evans et al., 2013). We run the temperature output from TraCE-21ka through Sedproxy, a forward proxy-system model for Mg/Ca and $U_{37}^{K'}$ sediment proxies (Dolman & Laepple, 2018; Laepple & Huybers, 2013). Sedproxy is capable of reproducing the spectral behavior of the different proxies (i.e., steeper spectral slopes in $U_{37}^{K'}$ records and more variance in Mg/Ca records (Laepple & Huybers, 2013)). In this study we do not aim for comprehensive assessment of the proxy-system model, or a full exploration of its parameter space. Rather, we use it to provide a general indication of how non-climatic processes may impact a proxy record's spectral characteristics.

Sedproxy is based on the framework for proxy-system modeling laid out in Evans et al. (2013), and is comprised of three parts: (a) a sensor model, which accounts for seasonal biases in how the sediment proxies (Mg/Ca and $U_{37}^{K'}$) respond to a temperature forcing (i.e., the tendency for these proxies to better record temperature during periods of proxy growth); (b) an archive model, which accounts for how the proxies are recorded within the sediment. In this case, the archive model is a model for post-depositional bioturbation, wherein the upper few centimeters of marine sediments are mixed by burrowing organisms, which acts to smooth and mix the climate signal on a range of timescales; and (c) an observation model, which accounts for potential errors or biases in how scientists measure and extract a climate signal from the sediment core. Proxies are recording temperature at approximately monthly resolution, while the resolution of temperature reconstructions are on the order of decades to centuries, and high-frequency variability can be aliased onto the low-frequency variability. Mg/Ca and $U_{37}^{K'}$ samples are also taken from layers that are 1–2 cm thick, which can smooth the temperature signal. Also included in the observation model are potential sample and measurement errors, for example, due to different instruments or cleaning techniques used by different labs (Dolman & Laepple, 2018).

However, Sedproxy does not simulate chronology errors, which could ultimately alter the coherence between proxy records at different sites. Temperature reconstructions using sediment cores rely on age-depth models, which predict age at any depth in the sediment core based on a small number of age estimates downcore. In the database compiled here, these age estimates primarily come from radiocarbon dating (Leduc et al., 2010b). We thus combine Sedproxy with an age-depth model adapted from Blaauw and Christeny (2011). For each site, we create a synthetic sedimentation record, where the mean sediment accumulation rate is μ_x (yrs per cm) and the standard deviation of the sediment accumulation is σ_x. The record is separated into n segments separated by Δc cm. For each segment $j = 1, 2, \ldots n$,

$$x_j = w x_{j+1} + (1 - w)\alpha_j, \tag{3}$$

where x_j is the sediment accumulation rate in segment j and w represents the "memory" in the accumulation rate. In this model, the sediment accumulation rate at depth c_j is a weighted average of the accumulation rate at the previous depth c_{j+1} and some independent noise α_j, where $w = 0$ would represent completely independent accumulation rates between segments and $w = 1$ would represent a fixed accumulation rate throughout the core. The noise term follows a Gamma distribution, $\alpha \sim Gamma(a_\alpha, b_\alpha)$, with $a_\alpha = \left(\mu_x/\sigma_x\right)^2$ and $b_\alpha = (\sigma_x)^2/\mu_x$. The distribution of w follows,

$$f(w) = \frac{d_s w^{\frac{d_s}{\Delta c}-1}}{\Delta c} f_R\left(w^{\frac{d_s}{\Delta c}}\right) \tag{4}$$

Here, R is the correlation between any two segments separated by d_s, whose distribution f_R is assumed to follow a Beta distribution $R \sim Beta(a_w, b_w)$. This produces a "true" synthetic age-depth record.

We then simulate an age-depth model. We "sample" the core at m discrete depths c_i, where $i = 1, 2, \ldots m$, and assign an associated age $\tilde{\phi}_i = \phi_i + \epsilon$. Here, ϕ_i is the true age at depth i, $\tilde{\phi}_i$ is the estimated ("jittered") age, and

Table 1
Parameter Values Used to Generate Pseudo-Proxies

Parameter	Value	Reference
No. of samples (Mg/Ca)	30	Dolman and Laepple (2018)
Measurement error	0.5°C (Mg/Ca), 0.25°C ($U_{37}^{K'}$)	Dolman and Laepple (2018)
Inter-individual variation in Mg/Ca proxies ("vital effects") (Sadekov et al., 2008; Haarmann et al., 2011)	2°C (Mg/Ca), 0°C ($U_{37}^{K'}$)	Dolman and Laepple (2018)
Mean sedimentation rate	50 cm/kyr	Based on mean sedimentation rates of Holocene cores in Leduc et al. (2010a), which vary from 2 to 213 cm/kyr
Standard deviation of sedimentation rate	100 cm/kyr	Not well-constrained, based on original sources in Leduc et al. (2010b)
No. of age-depth estimates throughout the column	10	Estimated based on original sources in Leduc et al. (2010b)

$\epsilon \sim N(\mu_\phi, \sigma_\phi)$ is a normally-distributed age error. We linearly interpolate between the "jittered" ages $\tilde{\phi}_i$ to generate age estimates at each depth in the core. This produces a linear age-depth model, where sedimentation rates are assumed to be constant between age-depth samples. This is the methodology used for many of the records in our database (Leduc et al., 2010a). An illustration of the age-depth model is shown in Supporting Information S1 (Figure S5).

Using Sedproxy and the synthetic age-depth model, we produce "pseudo-proxies" at each site. These pseudo-proxies reflect our best understanding of how the original temperature signals (in this case, TraCE-21ka) may be filtered by Mg/Ca and $U_{37}^{K'}$. We choose characteristic parameter values based on the literature (Table 1) to use at every site, although in reality parameter values are expected to vary by location (and potentially through time) (e.g., Dolman & Laepple, 2018). Forward-modeled pseudoproxies are in units of mmol/mol (Mg/Ca) and $U_{37}^{K'}$ index ($U_{37}^{K'}$), which are then transformed into sea-surface temperature values using the same calibrations used to standardize the proxy database: Müller et al. (1998) and Anand et al. (2003) for $U_{37}^{K'}$ and Mg/Ca, respectively. After generating pseudo-proxies at each site, we then estimate the frequency-dependent behavior of the pseudo-proxies, using the same methodology described in Section 2.3.

3. Proxy-Model Comparison

3.1. Spectral Slopes

A key feature of forced simulations is that the magnitude of low-frequency temperature variability is amplified relative to unforced simulations (e.g., Parsons et al., 2020). At decadal timescales, spectral slopes (β_{10}) are similar across all simulations (piControl, past1000, and TraCE-21ka, Figures 4a, 4d, and 4g). β_{10} are white or red everywhere except the equatorial Pacific, where blue slopes are associated with a reduction in energy at timescales longer than the high-energy El-Niño Southern Oscillation (e.g., Ault, Deser, et al., 2013). Similarities in β_{10} are most striking between piControl and past1000 simulations (Figures 4a and 4d), although TraCE-21ka is also largely similar albeit with overall more blue β_{10} (particularly in the North Atlantic and over Greenland; Figure 4g).

However, on timescales longer than a decade, distinctions begin to appear between unforced and forced simulations. On centennial timescales, spectral slopes (β_{100}) are steeper (redder) in the forced simulations (past1000 and TraCE-21ka cf. piControl; Figures 4e, 4h cf. Figure 4b). In this spectral band, β_{past1000} and β_{TraCE} are globally redder than $\beta_{\text{piControl}}$, and particularly in the tropics (past1000) and North Atlantic (TraCE-21ka). On the longest timescales (400 years–2 kyr), β_{TraCE} continue to steepen. While piControl simulations are not long enough to provide an estimate of unforced $\beta_{\text{piControl,1000}}$, the spectra of the piControl simulations flatten out on multidecadal to multicentennial timescales (Figure S4 in Supporting Information S1), suggesting that $\beta_{\text{piControl,1000}}$ may approach zero.

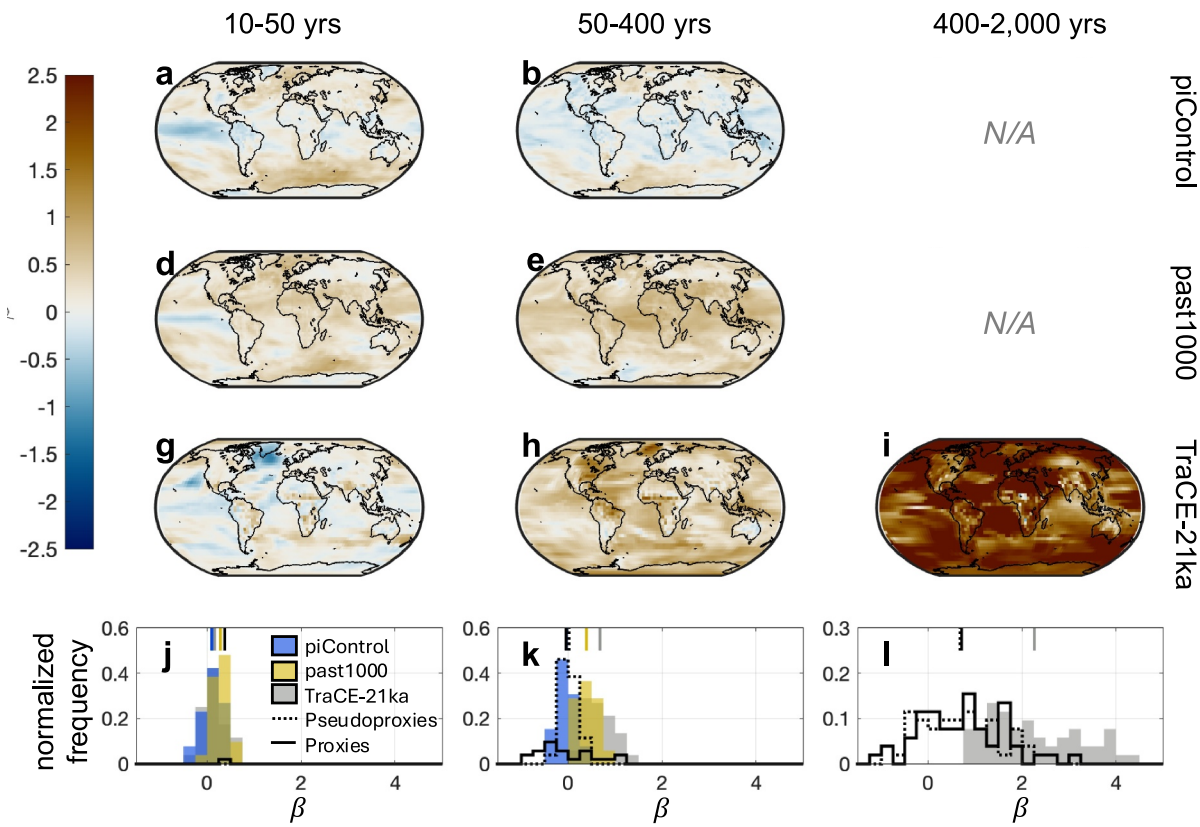

Figure 4. Spectral slope (β) estimates for models and proxy reconstructions. (a–i) Spectral slope estimations in each frequency bin at all gridpoints in preindustrial control (piControl, unforced; a–c) simulations, past-millennium simulations (past1000, forced; d–f), and TraCE-21k simulations (forced; g–i). For piControl and past1000, mean β across nine simulations is shown. No β estimates are available for piControl or past1000 in the 400 years–2 kyr spectral band. (j–l) Normalized histograms of spectral slope estimates for piControl simulations (blue), past1000 simulations (yellow), and TraCE-21k simulations (gray) at each gridpoint for which a proxy record exists. Spectral slopes estimated in proxy records are shown for comparison (black), along with spectral slopes in pseudo-proxies (forced with TraCE-21ka; dotted black). Colored tick-marks on the top *x*-axis indicate the mean value for each distribution.

The steepening spectral slopes in TraCE-21ka on these longer timescales could be due either to: (a) higher-frequency variability accumulating on long-period variations because of memory associated with the oceans (e.g., Huybers & Curry, 2006); or (b) the nature of the forcing in this simulation, which is primarily implemented as meltwater fluxes and ice-sheet coverage that vary on multicentennial timescales. Single-forcing experiments from TraCE-21ka indicate that most of the steepening can be attributed to the latter explanation (slowly-evolving forcings; see Figure S7 in Supporting Information S1). The overall spatial patterns in Figures 4a–4h are consistent with past studies, which have suggested that there are broad regional patterns of β that vary with frequency (e.g., Fredriksen & Rypdal, 2016; Henriksson et al., 2015; Huybers & Curry, 2006; Parsons et al., 2017).

We now isolate the spectral slopes at locations where proxy records are located, and these are presented as histograms in Figures 4j–4l. We see again the steeper spectral slopes in the forced simulations, particularly at lower frequencies (yellow [past1000] and gray [TraCE21-ka] cf. blue [piControl]). Evaluated statistically, the spectral-slope distributions in the unforced (piControl) and forced (past1000, TraCE-21ka) simulations are different at the 5% significance level at both decadal and centennial timescales, as determined by a Kolmogorov-Smirnov test (Massey, 1951). There is a clear separation in variability between forced and unforced simulations that more strongly emerges on long timescales. This behavior is consistent with previous work that found that the number of unique temperature patterns in forced and unforced simulations are similar at high frequencies, but not at lower frequencies (multidecadal timescales and longer): forcing adds spatially coherent variability on these longer timescales, ultimately amplifying low-frequency variability and reducing the independence in the system (Cleveland Stout et al., 2023).

Next, we estimate biases in β introduced by the proxy system itself. When TraCE-21ka temperature output is filtered by the proxy-system model, slopes generally whiten, biasing $\beta_{\text{pseudo}-\text{proxy}}$ low compared with β_{TraCE} on centennial-and-longer timescales (Figures 4k and 4l, dotted histogram [pseudo-proxies] cf. gray histogram [TraCE-21ka]). While bioturbation reduces high-frequency variability, other sources of error and bias (including measurement errors, individual variations between samples—which are sometimes called vital effects—and seasonal aliasing) amplify high-frequency variability and flatten (whiten) the spectrum (Dolman et al., 2021) [see Figure S6 in Supporting Information S1]. The discrepancy between $\beta_{\text{pseudo}-\text{proxy}}$ and β_{TraCE} is more prevalent on centennial timescales (Figure 4k cf. Figure 4l), because random errors and systematic biases that contribute to $\beta_{\text{pseudo}-\text{proxy}}$ primarily influence high-frequency variability and β_{TraCE} is thus better-preserved by the pseudo-proxy network on longer timescales (Dolman et al., 2021).

Finally, we look at β_{proxy}. On centennial timescales (i.e., β_{100}), the first thing we notice is that the spread in $\beta_{\text{proxy},100}$ is considerably larger than in any of the model outputs (Figure 4k, black distribution [proxies] cf. blue [piControl], yellow [past1000], and gray [TraCE-21ka] distributions). This spread can also be seen in Figure 4h, which shows large site-to-site variations in $\beta_{\text{proxy},100}$ (white slopes in close proximity to red slopes). We note that $\beta_{\text{proxy},100}$ have large uncertainties on this timescale (see Figure S2 in Supporting Information S1); at each site, even the sign of $\beta_{\text{proxy},100}$ is poorly constrained, which is consistent with a smaller signal-to-noise ratio at high frequencies (Dolman et al., 2021; Reschke et al., 2019). We see that $\beta_{\text{psuedo}-\text{proxy},100}$ has a tighter range than $\beta_{\text{proxy},100}$ on 50–400 years timescales (Figure 4k, black distribution [proxies] cf. dotted-black distribution [pseudo-proxies]), which may mean that our proxy-system model underestimates errors in this frequency band. Other proxies that are capable of resolving centennial and multicentennial variability may thus be more useful in constraining β_{100}. On millennial timescales (β_{1000}, Figure 4l), the range of $\beta_{\text{psuedo}-\text{proxy},1000}$ is consistent with the range of $\beta_{\text{proxy},1000}$ (Figure 4l, solid black distribution [proxies] cf. dotted-black distribution [pseudo-proxies]), which suggests our proxy-system model may be more appropriate at capturing errors on these longer timescales.

On this longest timescale (β_{1000}), the distribution of β_{proxy} shifts right (toward higher, redder values; Figure 4l), which is consistent with forced variability that amplifies low-frequency variability. We note that $\beta_{\text{proxy},1000}$ are consistently whiter (less steep) than $\beta_{\text{TraCE},1000}$ (solid black distribution [proxies] cf. gray distribution [TraCE-21ka]). However, the median value of $\beta_{\text{proxy},1000}$ is in close agreement with the median value of $\beta_{\text{pseudo}-\text{proxy},1000}$ (Figure 4l, solid black line cf. dotted-black line), because $\beta_{\text{pseudo}-\text{proxy},1000}$ is whitened relative to $\beta_{\text{TraCE},1000}$. If we take the implications of our proxy-system model at face value, it might suggest that the climate recorded by the proxies is consistent with the slopes of TraCE-21ka. Briefly, we also note that the spatial pattern of $\beta_{\text{proxy},1000}$ on 400 years–2 kyr timescales has smaller site-to-site variations in $\beta_{\text{proxy},1000}$ than on 50–400 years timescales (Figure 4h cf. Figure 4i), although they exceed site-to-site variations in $\beta_{\text{TraCE},1000}$.

3.2. Cross-Coherence

We aim to understand the spatial scale of temperature variability in the climate simulations and the proxy record —that is, whether variability is locally confined, regionally confined, or global. To that end, we compute cross-coherence pairwise between all sites following the methodology in Section 2.3. Coherence is averaged in each frequency band, producing a grid of coherence estimates that have been averaged across all models in a given experiment (Figure 5). Each square in the grid corresponds to a multi-model-mean coherence estimate between two sites, where sites are ordered by latitude following Figure 2. This ordering allows us to see which regions have similar temperature variability at different frequencies; more coherent regions are redder, while less coherent regions are whiter. We also show the percentage of sites for which coherence >0.6 (Figure 5, top-left quadrant). Note that this metric does not take account of the extra statistical information available from other adjacent sites, and thus is not representative of the spatial patterns of coherence. We use it here as a reference value to characterize the degree of coherence across the network. Diagonal values, which are all ones because a timeseries is perfectly coherent with itself, are shown as null (gray). Because the piControl and past1000 experiments have a maximum simulation time of 1,200 years, average coherence values for the last frequency band (400–2,000 years) are not shown.

Across model simulations, there are two robust features: (a) coherence increases toward longer timescales and (b) coherence is overall higher, and has more spatial structure (more, and larger, correlated regions of high coherence)

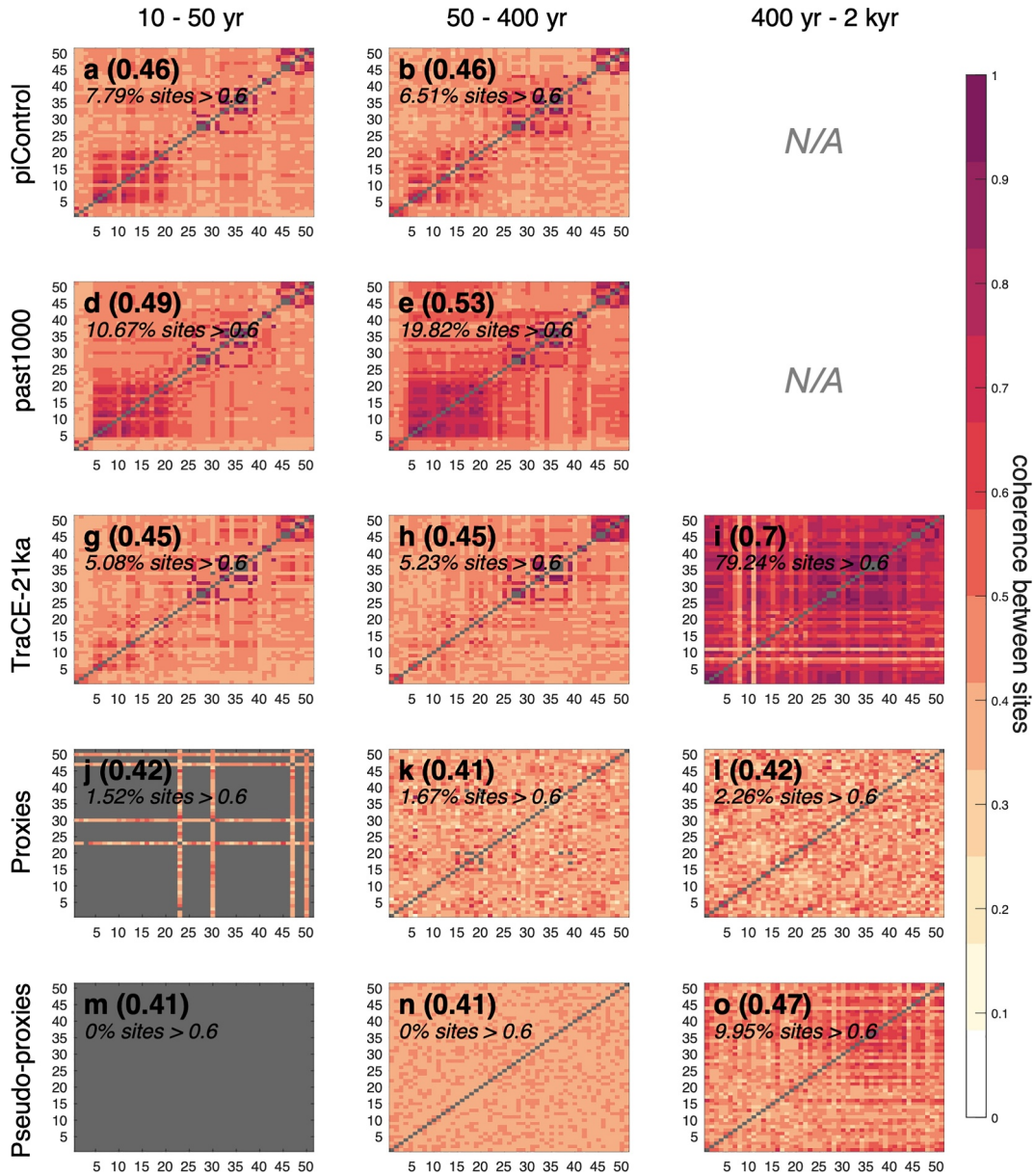

Figure 5. Pairwise cross-coherence in piControl simulations (a–c), past1000 simulations (d–f), TraCE-21ka simulations (g–i), pseudo-proxies forced by TraCE-21ka (j–l), and sea-surface temperature reconstructions from proxies (m–o). Gray indicates null values. Diagonal values are also shown as null. For model simulations, coherence between sites in the same grid cell are shown as null. Numbering on the *x*- and *y*-axis represents site number, consistent with Figure 1 and ordered from southernmost site (1) to northernmost site (48). Each gridcell represents the cross-coherence estimate between temperature at two sites averaged within a given frequency band. Mean pairwise coherence across all sites, excluding null values, are shown in the top left, as is the percentage of sites for which the mean cross-coherence is greater than 0.6 within a given frequency band. For piControl and past1000, coherence is the multi-model mean. No coherence estimates are available for piControl or past1000 in the 400 years–2 kyr spectral band.

in the forced simulations (past1000 and TraCE-21ka) compared with the unforced simulations (piControl). Cross-coherence among sites is relatively low across all model simulations on short timescales (10–50 years); ignoring the diagonals, mean coherence is 0.46 for piControl, 0.49 for past1000, and 0.45 for TraCE-21ka (Figures 5a, 5d, and 5g). On 50–400 years timescales, cross-coherence remains relatively low in piControl and TraCE-21ka simulations (mean coherence of 0.46 and 0.45, respectively; Figures 5b and 5h), but slightly increases in the past1000 simulations (mean coherence of 0.53; Figure 5e cf. Figure 5d). On the longest timescales (400–2 kyr),

cross-coherence dramatically increases in the forced simulation TraCE-21ka (mean value of 0.70; Figure 5i cf. Figures 5g and 5h). We interpret this as reflecting the prescribed freshwater forcing in TRaCE as having a global impact on ocean conditions on these longest timescales (see Figure S8 in Supporting Information S1).

We next discuss the spatial structure of the coherence (5 a-i). In all simulations and at all frequencies, cross-coherence is relatively high among sites in the Mediterranean (∼ sites 26, 32–33, and 35–37) because of their geographic proximity (Figure 1). In the past1000 simulations, coherence is high in the tropics (sites ∼5 to 22). This higher tropical coherence is missing in the TRaCE-21ka simulations, suggesting it is due to the volcanic forcing in past1000, which TRaCE-21ka lacks. In all simulations, coherence is high among the North Atlantic sites (44–52). It is highest in TRaCE-21ka, likely reflecting the importance of the ice-sheet and meltwater forcing applied in that simulation (also see Figure S8 in Supporting Information S1). Overall, these results point toward more coherent variability across proxy sites (i.e., across the marine mid- and low-latitudes), with a stronger spatial structure of cross-coherence, on multicenntennial-to-millennial timescales in the forced simulations compared with the unforced simulations.

Cross-coherence among proxy temperature records are slightly harder to characterize because some records do not overlap, and thus do not yield coherence estimates (null values shown as gray in Figures 5n and 5o). The mean cross-coherence is 0.41 at 50–400 years timescales, and is only slightly higher (0.42) at 400–2,000 years timescales. There are no spatial patterns that stick out; coherence is very noisy (Figures 5k and 5l), which is also consistent with noisy β estimates (Figures 4k and 4l). Some sites do exceed 0.6 coherence in both the 50–400 and 400 years–2 kyr frequency bands (1.67% and 2.26%, respectively). However, the sites exceeding 0.6 are not clustered together as they are in the simulations and pseudo-proxy experiment. Finally, the low temporal resolution of the proxy records, and the shorter periods where records overlap, strongly decreases the degrees of freedom when comparing two records, increasing the likelihood of generating high coherence values by chance. This suggests that individual values of high cross-coherence in Figures 5k and 5l may not be physically robust and be should not be over-interpreted.

There are three possible explanations for low, noisy cross-coherence in the proxy records. These possibilities are: (a) there are strong local dynamics that drive variability and that are not reflected in our models, resulting in incoherent variability even on regional scales and even on the longest (400–2,000 years) timescales; (b) there is substantial forced variability, but the proxy-to-temperature transfer function and age-depth model contain biases and errors that have destroyed much of the coherence signal; or (c) some combination of both. We briefly explore the second possibility by running a "pseudo-proxy experiment" and attempting to quantify how age uncertainties and Mg/Ca and $U_{37}^{K'}$ biases may influence predicted coherence in TraCE-21ka. The pseudo-proxy coherence is indeed reduced relative to the true (TraCE-21ka) coherence (means of 0.41 0.47 in the pseudo-proxies, cf. 0.45 and 0.70 in TraCE-21ka, for 50–400 years timescales and 400–2,000 years timescales respectively). However, some spatial structure is preserved—particularly by the $U_{37}^{K'}$ records (not shown)—at millennial timescales. In the North Atlantic, where coherence is relatively high in the TraCE-21ka simulations, the pseudo-proxies remain relatively coherent with each other (Figure 5i cf. Figure 5o, top-right quadrant). If there were some coherent regional patches of temperature variability, our pseudo-proxy experiment suggests that they would not be completely destroyed, although it is possible that our proxy-system model fails to account for some sources of bias or error that further degrade coherence signals, including spatial variations in parameter values.

4. Summary and Discussion

We have sought to evaluate whether the existing network of marine-sediment cores shows distinct spatial fingerprints of forced or unforced variability. To that end, we used a suite of forced and unforced models with long integration times (≥400 years) to evaluate the impact of forcing on frequency-dependent behavior; namely, (a) spectral slope β, which measures how variability changes from high frequencies to low frequencies and (b) coherence, which measures whether variability is global or isolated regionally or locally. Past studies have cataloged proxy-model discrepancies in the overall *magnitude* of temperature variability (e.g., Laepple et al., 2023; Laepple & Huybers, 2014b); here, we focus on the *spatial pattern of variability*, which may shed light on the physical mechanisms driving the proxy-model mismatch.

At inter-decadal and longer timescales, forced models exhibit significantly different (steeper) spectral slopes, which is consistent with large-scale forcings being integrated by the climate system and enhancing low-frequency

variability. Spectral slopes also vary relatively smoothly in space (i.e., adjacent sites have similar β). In forced simulations, we found that the spatial pattern of variability is also heavily dependent on the forcing mechanisms (e.g., simulations with volcanic eruptions have spatial patterns of cross-coherence and β that are distinct from simulations with evolving ice sheets). In our globally-distributed set of proxy temperature reconstructions, we found a wide spread in spectral slope, including some substantial differences between adjacent sites. Two key targets for future work are (a) dynamically constraining these site-to-site variations in β—that is, determining to what extent the physical climate system allows β to vary between adjacent sites and (b) understanding how biases introduced by the proxy recording mechanisms (e.g., timing uncertainty, bioturbation) alter the spectral characteristics of temperature proxies (following Dolman et al., 2021; Kunz et al., 2020).

When we filter the climate model output through a proxy-system model to account for proxy physical processes, the predicted pseudo-proxies suggest that proxies damp variability on long timescales, biasing β_{proxy} low (toward zero). At face value, this would imply the spectral slope of the temperature experienced by the proxies is higher than β_{proxy}, bringing the "true" climate spectrum more in-line with β_{TraCE}. The implication here is that TraCE-21ka may include the forcings that are relevant on these long, millennial timescales. High β—as is found in TraCE-21ka, and the bias-corrected sediment proxies—is consistent with low-frequency temperature variability with a large forced component.

Similarly, models indicate that the addition of forcing increases coherence on long timescales, particularly in regions where forcing is being applied. Variability at tropical sites is sensitive to volcanic forcing, and variability at high-latitude sites is sensitive to ice-sheet and meltwater forcing. Proxy records, by contrast, exhibit very little coherence, suggesting that either significant unforced variability exists or that the coherence was destroyed by biases and noise introduced by the proxy system. We probed the latter hypothesis using a proxy-system model, and find that, for parameter values typical of Holocene Mg/Ca and $U^{K'}_{37}$ proxies, millennial-scale coherence should be detectable in a proxy record if it once existed in the climate. However, it is also possible our proxy-system model does not capture all processes that could corrupt the coherence signal in the proxy record. If our proxy-system model does capture all such processes, our results would suggest that the mismatch between proxies and climate models in low-frequency variability is due to an underestimation of regional unforced low-frequency variability in GCMs.

Our analyses thus reveal a somewhat mixed picture, with significant inconsistencies between models and proxy reconstructions that cannot be easily reconciled. Steeper spectral slopes in proxies point to dominant external forcing, in-line with TraCE-21ka, but low spatial coherence among proxies compared to models implies large localized internal variability. Some of the proxy-model mismatch in β and coherence may be related to differences in spatial resolution; the models here do not explicitly simulate sub-mesoscale variability, which is what proxies record, although past work has pointed toward the existence of a proxy-model mismatch in variability even on larger, regional spatial scales (e.g., Bothe et al., 2015; Parsons et al., 2017). It is our view that the low spatial coherence between nearby cores is a cause for concern about the degree to which the cores record large-scale climate information, including the overall direction of temperature change over this period. The lack of spatial coherence between proxy sites relative to models–even on long, multi-centennial timescales–may also have implications for paleo data assimilation products, which use model-constrained spatial correlations to incorporate information from proxy records.

The ability to resolve some of these inconsistencies is somewhat hampered by the configuration of the current sedimentary core network. One specific example can be seen in Figure 5. Although the specific patterns in Figures 5b, 5e, and 5h are an artifact of the differences in the way the model simulations were forced, as currently constituted, the network struggles to distinguish among them, at least on the timescales that past1000 simulations can resolve (Figures 5k and 5l). One can ask: what hypothetical proxy network could, in principle, distinguish between these forced (past1000, TraCE-21ka) and unforced patterns? Key aspects of such a network design would include temporal resolution; reproducibility of adjacent cores; and reducing the non-climatic errors introduced into the proxies. Our analysis framework could be used to address that question through a fuller exploration of the potential parameter space within the proxy-system framework. The analysis presented here suggests that it may be particularly important to gather cores in regions sensitive to forcing; if sites within those regions are coherent with each other, that would suggest forced variability; on the other hand, if sites within those regions are incoherent with each other, that would suggest unforced variability. If there are no regions in which sites are coherent with each other, this may suggest that proxies are too noisy to reliably constrain variability,

unless there are plausible dynamical explanations (e.g., weaker diffusion or inadequate representation of sub-grid-scale processes Laepple et al., 2023) that might produce high-magnitude localized variability.

Here, we chose to analyse sediment cores to make progress on understanding temperature variability on centennial-to-millennial timescales. These cores have maximum resolvable frequencies between 1/8 and 1/400 years, depending on the site, with a mean resolvable frequency of 1/97 years. This is at the low end of the frequencies that can be resolved by available piControl and past1000 simulations. It is thus of high priority to generate a larger set of longer simulations that can better constrain the behavior of low-frequency variability. Such simulations should include a range of forcing mechanisms (no forcing, episodic volcanic forcing imposed in specific regions, and slowly-evolving forcings like meltwater fluxes and ice-sheet changes). There are also other proxies available whose timescales better overlap with existing model simulations—our analyses might be applied to understand decadal- and centennial-scale variability using temperature reconstructions from tree rings and corals that have higher temporal resolution and more accurate chronologies, or online data assimilation products that offer physically-constrained temperature estimates with high temporal resolutions.

Our analyses shed further light on the discrepancies between climate model output and proxy-temperature reconstructions. Broadly, our results point toward the importance of considering the relationships among proxy records. Climate reconstructions from proxy networks should, in principle, be dynamically consistent: coherence and spectral slopes should increase toward lower frequencies and, under forcing, there should be (a) steep spectral slopes and (b) strongly coherent temperature signals on long timescales, especially in the regions where the forcing is applied. We propose two directions for future work (a) characterizing statistical fingerprints of different *individual* forcing mechanisms, such as we briefly do with the single-forcing Trace-21ka simulations. (b) Continued work on proxy-system models, including models that contain stochastic terms and time-scale dependence, to better account for how proxies alter the underlying climate signal.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The compiled database can be found at Zenodo (Cleveland Stout, 2025). Model output is available through the NCAR Research Data Archive (B. Otto-Bliesner & Rosenbloom, 2021) (TraCE-21ka) and https://aims2.llnl.gov/search (CMIP5; preindustrial control and last millennium simulations). The proxy-system model Sedproxy is available in R through CRAN (Laepple & Dolman, 2022). Our implementation of Sedproxy in MATLAB, with an age-depth model, is stored on Zenodo (Cleveland Stout, 2025).

Acknowledgments
This research was funded by a Schlanger Ocean Drilling Fellowship through the U.S. Science Support Program (RCS), NSF Grant P2C2-2102829 (CP, GR, and RCS), and AGS-2019647 (GR).

References

Anand, P., Elderfield, H., & Conte, M. H. (2003). Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography*, *18*(2). https://doi.org/10.1029/2002pa000846

Askjær, T. G., Zhang, Q., Schenk, F., Ljungqvist, F. C., Lu, Z., Brierley, C. M., et al. (2022). Multi-centennial Holocene climate variability in proxy records and transient model simulations. *Quaternary Science Reviews*, *296*, 107801. https://doi.org/10.1016/j.quascirev.2022.107801

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., George, S. S., Otto-Bliesner, B. L., et al. (2013). The Continuum of hydroclimate variability in Western North America during the last millennium. *Journal of Climate*, *26*(16), 5863–5878. https://doi.org/10.1175/JCLI-D-11-00732.1

Ault, T. R., Deser, C., Newman, M., & Emile-Geay, J. (2013). Characterizing decadal to centennial variability in the equatorial Pacific during the last millennium. *Geophysical Research Letters*, *40*(13), 3450–3456. https://doi.org/10.1002/grl.50647

Bindoff, N., Stott, P. A., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D., et al. (2013). Detection and attribution of climate change: From global to regional. In T. 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*.

Blaauw, M., & Christeny, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis*, *6*(3), 457–474. https://doi.org/10.1214/11-BA618

Bothe, O., Evans, M. N., Donado, L. F., Bustamante, E. G., Gergis, J., Gonzalez-Rouco, J. F., & Zorita, E. (2015). Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Climate of the Past*, *11*(12), 1673–1699. https://doi.org/10.5194/cp-11-1673-2015

Briner, J. P., McKay, N. P., Axford, Y., Bennike, O., Bradley, R. S., de Vernal, A., et al. (2016). Holocene climate change in Arctic Canada and Greenland. *Quaternary Science Reviews*, *147*, 340–364. https://doi.org/10.1016/j.quascirev.2016.02.010

Brown, P. T., Li, W., Li, L., & Ming, Y. (2014). Top-of-atmosphere radiative contribution to unforced decadal global temperature variability in climate models. *Geophysical Research Letters*, *41*(14), 5175–5183. https://doi.org/10.1002/2014GL060625

Brown, P. T., Li, W., & Xie, S. P. (2015). Regions of significant influence on unforced global mean surface air temperature variability in climate models. *Journal of Geophysical Research: Atmospheres*, *120*, 480–494. https://doi.org/10.1038/175238c0

Català, A., Cacho, I., Frigola, J., Pena, L. D., & Lirer, F. (2019). Holocene hydrography evolution in the Alboran Sea: A multi-record and multi-proxy comparison. *Climate of the Past*, *15*(3), 927–942. https://doi.org/10.5194/cp-15-927-2019

Cleveland Stout, R. (2025). Age-depth model [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.16787058

Cleveland Stout, R., Proistosescu, C., & Roe, G. (2023). Fingerprinting low-frequency last millennium temperature variability in forced and unforced climate models. *Journal of Climate*, *36*(20), 7005–7023. https://doi.org/10.1175/jcli-d-22-0810.1

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In T. F. Stocker, et al. (Eds.), *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the intergovernmental Panel on Climate Change* (Vol. 9781107057, pp. 1029–1136). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.024

Collins, M., Tett, S. F. B., & 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

Collins, W. D., Bitz, C. M., Blackmon, M., Bonan, G., Bretherton, C. S., Carton, J., et al. (2006). The community climate system model version 3 (CCSM3). *Journal of Climate*, *19*(11), 2122–2143. https://doi.org/10.1175/jcli3761.1

Conte, M. H., Sicre, M. A., Rühlemann, C., Weber, J. C., Schulte, S., Schulz-Bull, D., & Blanz, T. (2006). Global temperature calibration of the alkenone unsaturation index (Uk'37) in surface waters and comparison with surface sediments. *Geochemistry, Geophysics, Geosystems*, *7*(2). https://doi.org/10.1029/2005GC001054

Dee, S. G., Emile-Geay, J., Evans, M., Allam, A., Steig, E., & Thompson, D. (2015). PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems. *Journal of Advances in Modeling Earth Systems*, *7*(3), 1220–1247. https://doi.org/10.1002/2015MS000447

Dee, S. G., Parsons, L. A., Loope, G. R., Overpeck, J. T., Ault, T. R., & Emile-Geay, J. (2017). Improved spectral comparisons of paleoclimate models and observations via proxy system modeling: Implications for multi-decadal variability. *Earth and Planetary Science Letters*, *476*, 34–46. https://doi.org/10.1016/j.epsl.2017.07.036

Dolman, A. M., Kunz, T., Groeneveld, J., & Laepple, T. (2021). A spectral approach to estimating the timescale-dependent uncertainty of paleoclimate records—Part 2: Application and interpretation. *Climate of the Past*, *17*(2), 825–841. https://doi.org/10.5194/cp-17-825-2021

Dolman, A. M., & Laepple, T. (2018). Sedproxy: A forward model for sediment-archived climate proxies. *Climate of the Past*, *14*(12), 1851–1868. https://doi.org/10.5194/cp-14-1851-2018

Evans, M., Tolwinski-Ward, S. E., Thompson, D. M., & Anchukaitis, K. J. (2013). Applications of proxy system modeling in high resolution paleoclimatology. *Quaternary Science Reviews*, *76*, 16–28. https://doi.org/10.1016/j.quascirev.2013.05.024

Foster, G., & Rahmstorf, S. (2011). Global temperature evolution 1979–2010. *Environmental Research Letters*, *6*(4), 44022. https://doi.org/10.1088/1748-9326/6/4/044022

Fraedrich, K., Luksch, U., & Blender, R. (2004). 1/f model for long-time memory of the ocean surface temperature. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *70*(3), 4. https://doi.org/10.1103/PhysRevE.70.037301

Franzke, C. L. E., Barbosa, S., Blender, R., Fredriksen, H. B., Laepple, T., Lambert, F., et al. (2020). The structure of climate variability across scales. *Reviews of Geophysics*, *58*(2), e2019RG000657. 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

Haarmann, T., Hathorne, E. C., Mohtadi, M., Groeneveld, J., Klling, M., & Bickert, T. (2011). Mg/Ca ratios of single planktonic foraminifer shells and the potential to reconstruct the thermal seasonality of the water column. *Paleoceanography*, *26*(3), 1–14. https://doi.org/10.1029/2010PA002091

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., et al. (2016). The last millennium climate reanalysis project: Framework and first results. *Journal of Geophysical Research: Atmospheres*, *121*(12), 6745–6764. https://doi.org/10.1002/2016JD024751

Hasselmann, K. (1976). Stochastic climate models Part I. Theory. *Tellus*, *28*(6), 473–485. https://doi.org/10.3402/tellusa.v28i6.11316

Hausfather, Z., Drake, H. F., Abbott, T., & Schmidt, G. A. (2020). Evaluating the performance of past climate model projections. *Geophysical Research Letters*, *47*(1), 1–10. https://doi.org/10.1029/2019GL085378

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107. https://doi.org/10.1175/2009BAMS2607.1

He, F. (2011). Simulating transient climate evolution of the Last Deglaciation with CCSM3 (Doctoral dissertation). Retrieved from https://www.cgd.ucar.edu/sites/default/files/trace/He_PhD_dissertation_UW_2011.pdf

Henriksson, S. V., Räisänen, P., Silen, J., Järvinen, H., & Laaksonen, A. (2015). Improved power-law estimates from multiple samples provided by millennium climate simulations. *Theoretical and Applied Climatology*, *119*(3–4), 667–677. https://doi.org/10.1007/s00704-014-1132-0

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, *116*(1), 770–799. https://doi.org/10.1061/taceat.0006518

Huybers, P. J., & Curry, W. (2006). Links between annual, Milankovitch and continuum temperature variability. *Nature*, *441*(7091), 329–332. https://doi.org/10.1038/nature04745

Jones, G. S., Stott, P. A., & Christidis, N. (2013). Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, *118*(10), 4001–4024. https://doi.org/10.1002/jgrd.50239

Judd, E. J., Bhattacharya, T., & Ivany, L. C. (2020). A dynamical framework for interpreting Ancient Sea surface temperatures. *Geophysical Research Letters*, *47*(15), e2020GL089044. https://doi.org/10.1029/2020GL089044

Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2023). CHELSA-TraCE21k—High-resolution (1 km) downscaled transient temperature and precipitation data since the Last Glacial Maximum. *Climate of the Past*, *19*(2), 439–456. https://doi.org/10.5194/cp-19-439-2023

Kaufman, D., McKay, N., Routson, C., Erb, M., Dätwyler, C., Sommer, P. S., et al. (2020). Holocene global mean surface temperature, a multi-method reconstruction approach. *Scientific Data*, *7*(1), 1–13. https://doi.org/10.1038/s41597-020-0530-7

Kim, J. H., Rimbu, N., Lorenz, S. J., Lohmann, G., Nam, S. I., Schouten, S., et al. (2004). North Pacific and North Atlantic sea-surface temperature variability during the Holocene. *Quaternary Science Reviews*, *23*(20–22), 2141–2154. https://doi.org/10.1016/j.quascirev.2004.08.010

Kunz, T., Dolman, A. M., & Laepple, T. (2020). A spectral approach to estimating the timescale-dependent uncertainty of paleoclimate records—Part 1: Theoretical concept. *Climate of the Past*, *16*, 1469–1492. https://doi.org/10.5194/cp-16-1469-202

Laepple, T., & Dolman, A. (2022). sedproxy: Simulation of sediment archived climate proxy records [Software]. *CRAN: Contributed Packages. The R Foundation*. https://doi.org/10.32614/cran.package.sedproxy

Laepple, T., & Huybers, P. J. (2013). Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability. *Earth and Planetary Science Letters*, *375*, 418–429. https://doi.org/10.1016/j.epsl.2013.06.006

Laepple, T., & Huybers, P. J. (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. J. (2014b). Ocean surface temperature variability: Large model-data differences at decadal and longer periods. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(47), 16682–16687. https://doi.org/10.1073/pnas.1412077111

Laepple, T., Ziegler, E., Weitzel, N., Hébert, R., Ellerhoff, B., Schoch, P., et al. (2023). Regional but not global temperature variability underestimated by climate models at supradecadal timescales. *Nature Geoscience*, *16*(11), 958–966. https://doi.org/10.1038/s41561-023-01299-9

Large, W. G., & Danabasoglu, G. (2006). Attribution and impacts of upper-ocean biases in CCSM3. *Journal of Climate*, *19*(11), 2325–2346. https://doi.org/10.1175/JCLI3740.1

Leduc, G., Schneider, R. R., Kim, J., & Lohmann, G. (2010a). Holocene and Eemian Sea surface temperature trends as revealed by alkenone and Mg/Ca paleothermometry. *Quaternary Science Reviews*, *29*(7–8), 989–1004. https://doi.org/10.1016/j.quascirev.2010.01.004

Leduc, G., Schneider, R. R., Kim, J.-H., & Lohmann, G. (2010b). Mg/Ca-derived Holocene sea surface temperature database. *PANGAEA*. https://doi.org/10.1594/PANGAEA.736946

Leith, C. E. (1978). Predictability of climate. *Nature*, *276*(5686), 352–355. https://doi.org/10.1038/276352a0

Lerch, A. (2012). *An introduction to audio content analysis applications in signal processing and music informatics*. IEEE Press. Piscataway, NJ.

Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., et al. (2014). The Holocene temperature conundrum. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(34), E3501–E3505. https://doi.org/10.1073/pnas.1407229111

Ma, J., Xu, S., & Wang, B. (2019). Warm bias of sea surface temperature in Eastern boundary current regions—A study of effects of horizontal resolution in CESM. *Ocean Dynamics*, *69*(8), 939–954. https://doi.org/10.1007/s10236-019-01280-4

Mann, M. E., & Park, J. (1993). Spatial correlations of interdecadal variation in global surface temperatures. *Geophysical Research Letters*, *20*(11), 1055–1058. https://doi.org/10.1029/93gl00752

Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68–78. https://doi.org/10.2307/2280095

Meehl, G. A., Hu, A., Arblaster, J. M., Fasullo, J., & Trenberth, K. E. (2013). Externally forced and internally generated decadal climate variability associated with the interdecadal Pacific Oscillation. *Journal of Climate*, *26*(18), 7298–7310. https://doi.org/10.1175/JCLI-D-12-00548.1

Min, S. K., Legutke, S., Hense, A., & Kwon, W. T. (2005). Internal variability in a 1000-yr control simulation with the coupled climate model ECHO-G -1. Near-surface temperature, precipitation and mean sea level pressure. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *57*(4), 605–621. https://doi.org/10.3402/tellusa.v57i4.14712

Müller, P. J., Kirst, G., Ruhland, G., Von Storch, I., & Rosell-Melé, A. (1998). Calibration of the alkenone paleotemperature index U37K based on core-tops from the eastern South Atlantic and the global ocean (60°N–60°S). *Geochimica et Cosmochimica Acta*, *62*(10), 1757–1772. https://doi.org/10.1016/S0016-7037(98)00097-0

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M., et al. (2019). Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nature Geoscience*, *12*(8), 643–649. https://doi.org/10.1038/s41561-019-0400-0

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

Nürnberg, D., Bijma, J., & Hemleben, C. (1996). Assessing the reliability of magnesium in foraminiferal calcite as a proxy for water mass temperatures. *Geochimica et Cosmochimica Acta*, *60*(5), 803–814. https://doi.org/10.1016/0016-7037(95)00446-7

O'Brien, J. P., & Deser, C. (2023). Quantifying and understanding forced changes to unforced modes of atmospheric circulation variability over the North Pacific in a coupled model large ensemble. *Journal of Climate*, *36*(1), 19–37. https://doi.org/10.1175/JCLI-D-22-0101.1

Otto-Bliesner, B., & Rosenbloom, N. (2021). Simulation of the transient climate of the last 21,000 years (TraCE-21ka) [Dataset]. *Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory*. https://doi.org/10.5065/CXB5-TV56

Otto-Bliesner, B. L., Schneider, R., Brady, E. C., Kucera, M., Abe-Ouchi, A., Bard, E., et al. (2009). A comparison of PMIP2 model simulations and the MARGO proxy reconstruction for tropical sea surface temperatures at last glacial maximum. *Climate Dynamics*, *32*(6), 799–815. https://doi.org/10.1007/s00382-008-0509-0

PAGES2k Consortium, McKay, N. P., Kaufman, D. S., von Gunten, L., Wang, J., Anchukaitis, K. J., et al. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*(1), 170088. https://doi.org/10.1038/sdata.2017.88

Palmer, M. D., & McNeall, D. J. (2014). Internal variability of Earth's energy budget simulated by CMIP5 climate models. *Environmental Research Letters*, *9*(3), 034016. https://doi.org/10.1088/1748-9326/9/3/034016

Parsons, L. A., Brennan, M. K., Wills, R. C., & Proistosescu, C. (2020). Magnitudes and spatial patterns of interdecadal temperature variability in CMIP6. *Geophysical Research Letters*, *47*(7), 1–11. https://doi.org/10.1029/2019GL086588

Parsons, L. A., & Hakim, G. J. (2019). Local regions associated with interdecadal global temperature variability in the last millennium reanalysis and CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *124*(17–18), 9905–9917. https://doi.org/10.1029/2019JD030426

Parsons, L. A., Loope, G. R., Overpeck, J. T., Ault, T. R., Stouffer, R. J., & Cole, J. E. (2017). Temperature and precipitation variance in CMIP5 simulations and paleoclimate records of the last millennium. *Journal of Climate*, *30*(22), 8885–8912. https://doi.org/10.1175/JCLI-D-16-0863.1

Pelletier, J. D., & Turcotte, D. L. (1997). Long-range persistence in climatological and hydrological time series: Analysis, modeling and application to drought hazard assessment. *Journal of Hydrology*, *203*(1–4), 198–208. https://doi.org/10.1016/S0022-1694(97)00102-9

Prahl, F. G., & Wakeham, S. G. (1987). Calibration of unsaturation patterns in long-chain ketone compositions for palaeotemperature assessment. *Nature*, *330*, 367–369. https://doi.org/10.1038/330367a0

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*(7), 1–7. https://doi.org/10.1126/sciadv.1602821

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

Reschke, M., Kröner, I., & Laepple, T. (2021). Testing the consistency of Holocene and Last Glacial Maximum spatial correlations in temperature proxy records. *Journal of Quaternary Science*, *36*(1), 20–28. https://doi.org/10.1002/jqs.3245

Reschke, M., Rehfeld, K., & Laepple, T. (2019). Empirical estimate of the signal content of Holocene temperature proxy records. *Climate of the Past*, *15*(2), 521–537. https://doi.org/10.5194/cp-15-521-2019

Rhines, A., & Huybers, P. J. (2011). Estimation of spectral power laws in time uncertain series of data with application to the Greenland ice sheet project 2 $\delta^{18}O$ record. *Journal of Geophysical Research*, *116*(1), 1–9. https://doi.org/10.1029/2010JD014764

Rugenstein, M., Bloch-Johnson, J., Gregory, J. M., Andrews, T., Mauritsen, T., Li, C., et al. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, *47*(4), e2019GL083898. https://doi.org/10.1029/2019GL083898

Sadekov, A., Eggins, S. M., De Deckker, P., & Kroon, D. (2008). Uncertainties in seawater thermometry deriving from intratest and intertest Mg/Ca variability in Globigerinoides ruber. *Paleoceanography*, *23*(1), 1–12. https://doi.org/10.1029/2007PA001452

Sicre, M. A., Jalali, B., Eiríksson, J., Knudsen, K. L., Klein, V., & Pellichero, V. (2021). Trends and centennial-scale variability of surface water temperatures in the North Atlantic during the Holocene. *Quaternary Science Reviews*, *265*, 107033. https://doi.org/10.1016/j.quascirev.2021.107033

Solomon, S., Daniel, J. S., Neely, R. R., Vernier, J. P., Dutton, E. G., & Thomason, L. W. (2011). The persistently variable "background" stratospheric aerosol layer and global climate change. *Science*, *333*(6044), 866–870. https://doi.org/10.1126/science.1206027

Springford, A., Eadie, G. M., & Thomson, D. J. (2020). Improving the Lomb–Scargle Periodogram with the Thomson Multitaper. *The Astronomical Journal*, *159*(5), 205. https://doi.org/10.3847/1538-3881/ab7fa1

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., et al. (2019). Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. *Climate of the Past*, *15*(4), 1251–1273. https://doi.org/10.5194/cp-15-1251-2019

Thompson, A. J., Zhu, J., Poulsen, C. J., Tierney, J. E., & Skinner, C. B. (2022). Northern Hemisphere vegetation change drives a Holocene thermal maximum. *Science Advances*, *8*(15), 2–11. https://doi.org/10.1126/sciadv.abj6535

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, *70*(9), 1055–1096. https://doi.org/10.1109/proc.1982.12433

Tierney, J. E., Malevich, S. B., Gray, W. R., Vetter, L., & Thirumalai, K. (2019). Bayesian calibration of the Mg/Ca paleothermometer in planktic foraminifera. *Paleoceanography and Paleoclimatology*, *34*(12), 2005–2030. https://doi.org/10.1029/2019PA003744

Tierney, J. E., & Tingley, M. P. (2018). BAYSPLINE: A new calibration for the alkenone paleothermometer. *Paleoceanography and Paleoclimatology*, *33*(3), 281–301. https://doi.org/10.1002/2017PA003201

von der Heydt, A. S., Ashwin, P., Camp, C. D., Crucifix, M., Dijkstra, H. A., Ditlevsen, P., & Lenton, T. M. (2021). Quantification and interpretation of the climate variability record. *Global and Planetary Change*, *197*(2020), 103399. https://doi.org/10.1016/j.gloplacha.2020.103399

Wills, R. C., Battisti, D. S., Armour, K. C., Proistosescu, C., & Parsons, L. A. (2021). Slow modes of global temperature variability and their impact on climate sensitivity estimates. *Journal of Climate*, *34*(21), 8717–8738. https://doi.org/10.1175/jcli-d-20-1013.1

Wilson, P. S., Tomsett, A. C., & Toumi, R. (2003). Long-memory analysis of time series with missing values. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *68*(1), 4. https://doi.org/10.1103/PhysRevE.68.017103

Xu, G., Chang, P., Ramachandran, S., Danabasoglu, G., Yeager, S., Small, J., et al. (2022). Impacts of model horizontal resolution on mean sea surface temperature biases in the community Earth system model. *Journal of Geophysical Research: Oceans*, *127*(12), e2022JC019065. https://doi.org/10.1029/2022JC019065

Zhang, Q., Berntell, E., Axelsson, J., Chen, J., Han, Z., De Nooijer, W., et al. (2021). Simulating the mid-Holocene, last interglacial and mid-pliocene climate with ec-earth3-lr. *Geoscientific Model Development*, *14*(2), 1147–1169. https://doi.org/10.5194/gmd-14-1147-2021

Zhu, F., Emile-Geay, J., McKay, N. P., Hakim, G. J., Khider, D., Ault, T. R., et al. (2019). Climate models can correctly simulate the continuum of global-average temperature variability. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(18), 8728–8733. https://doi.org/10.1073/pnas.1809959116