# Fingerprinting Low-Frequency Last Millennium Temperature Variability in Forced and Unforced Climate Models

REBECCA CLEVELAND STOUT,[a] CRISTIAN PROISTOSESCU,[b] AND GERARD ROE[c]

[a] *Department of Atmospheric Sciences, University of Washington, Seattle, Washington*
[b] *Department of Atmospheric Sciences and Department of Geology, University of Illinois Urbana–Champaign, Urbana, Illinois*
[c] *Department of Earth and Space Sciences, University of Washington, Seattle, Washington*

(Manuscript received 28 October 2022, in final form 10 May 2023, accepted 22 May 2023)

ABSTRACT: Constraining unforced and forced climate variability impacts interpretations of past climate variations and predictions of future warming. However, comparing general circulation models (GCMs) and last millennium Holocene hydroclimate proxies reveals significant mismatches between simulated and reconstructed low-frequency variability at multidecadal and longer time scales. This mismatch suggests that existing simulations underestimate either external or internal drivers of climate variability. In addition, large differences arise across GCMs in both the magnitude and spatial pattern of low-frequency climate variability. Dynamical understanding of forced and unforced variability is expected to contribute to improved interpretations of paleoclimate variability. To that end, we develop a framework for fingerprinting spatiotemporal patterns of temperature variability in forced and unforced simulations. This framework relies on two frequency-dependent metrics: 1) degrees of freedom ($\equiv N$) and 2) spatial coherence. First, we use $N$ and spatial coherence to characterize variability across a suite of both preindustrial control (unforced) and last-millennium (forced) GCM simulations. Overall, we find that, at low frequencies and when forcings are added, regional independence in the climate system decreases, reflected in fewer $N$ and higher coherence between local and global mean surface temperature. We then present a simple three-box moist-static-energy-balance model for temperature variability, which is able to emulate key frequency-dependent behavior in the GCMs. This suggests that temperature variability in the GCM ensemble can be understood through Earth's energy budget and downgradient energy transport, and allows us to identify sources of polar-amplified variability. Finally, we discuss insights the three-box model can provide into model-to-model GCM differences.

SIGNIFICANCE STATEMENT: Forced and unforced temperature variability are poorly constrained and understood, particularly that at time scales longer than a decade. Here, we identify key differences in the time scale–dependent behavior of forced and unforced temperature variability using a combination of numerical climate models and principles of downgradient energy transport. This work, and the spatiotemporal characterizations of forced and unforced temperature variability that we generate, will aid in interpretations of proxy-based paleoclimate reconstructions and improve mechanistic understanding of variability.

KEYWORDS: Energy transport; Climate models; Numerical analysis/modeling; Climate variability; Interdecadal variability

## 1. Introduction

Climate variations can be either forced or unforced. Forced variability results from top-of-atmosphere energy imbalances due to external factors (e.g., changes in greenhouse gases, solar radiation, or volcanic aerosols), while unforced variability results from internal dynamics (e.g., ocean–atmosphere heat exchange or chaotic system dynamics; e.g., Kay et al. 2015). Our confidence in projections of future climate depends on our ability to disentangle the relative magnitude of forced and unforced variability in Earth's recent climate history, particularly on local to regional spatial scales (e.g., Hegerl et al. 2007). Traditionally, sources of variability are disentangled using general circulation models (GCMs): forced climate variability can be estimated through an ensemble mean across a large number of climate realizations, while unforced climate variability can be characterized from long control integrations where no external forcing is applied (e.g., Kay et al. 2015). On fast (subdecadal) time scales, GCM simulations of the recent past (from 1970 to present) forced by aerosol and greenhouse gas concentrations broadly agree with each other and with instrumental observations (Hausfather et al. 2020), and point toward El Niño–Southern Oscillation as the single largest source of global mean surface temperature variability (e.g., Brown et al. 2015). However, on longer (interdecadal) time scales, there is less agreement on the sources and magnitudes of temperature variability among GCMs (e.g., Brown et al. 2015; Parsons et al. 2020). There is also evidence that GCMs show less low-frequency regional variability than do proxies (Laepple and Huybers 2014; Bothe et al. 2015) and historical observations (DelSole 2006), although high uncertainties muddy comparisons between reconstructed and modeled temperature (Bakker et al. 2022).

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-22-0810.s1.

*Corresponding author*: Rebecca Cleveland Stout, rrcs@uw.edu

DOI: 10.1175/JCLI-D-22-0810.1

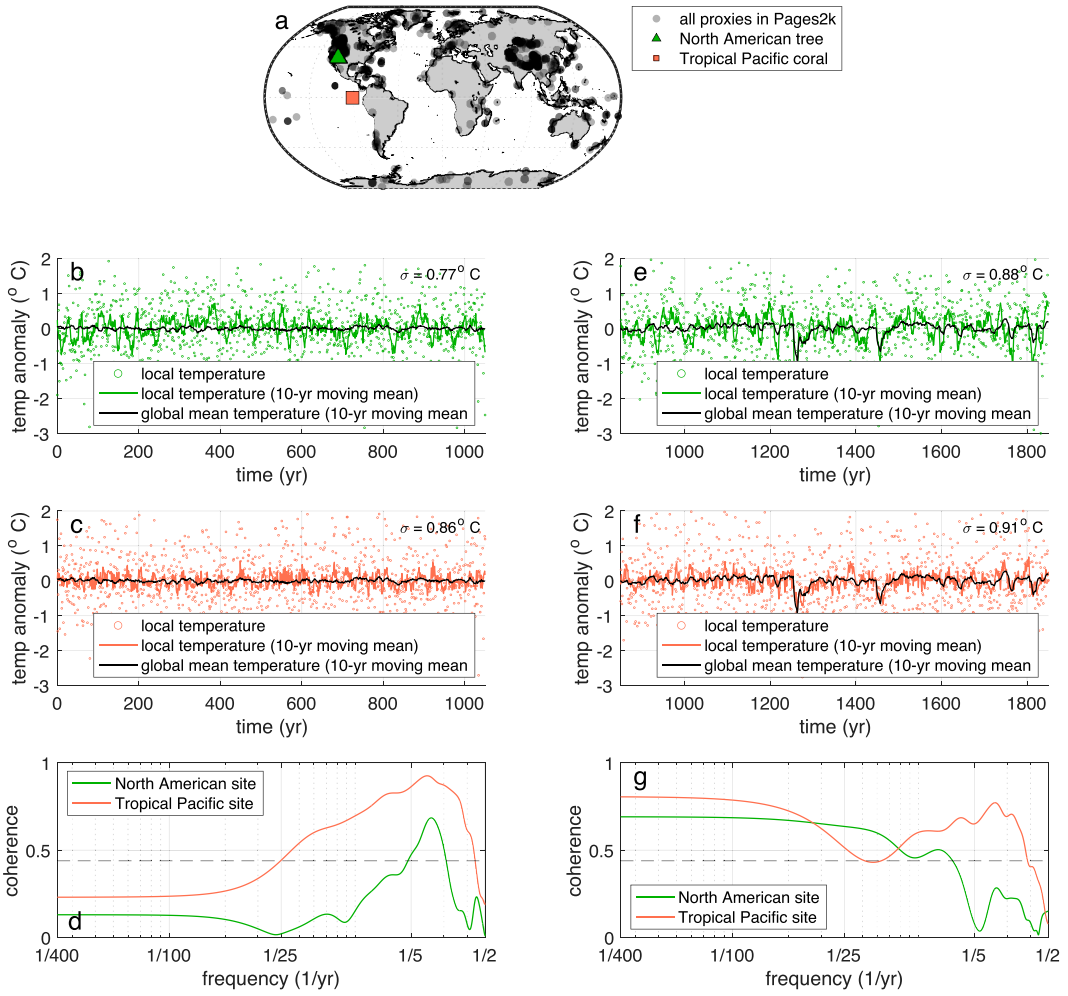


FIG. 1. Temperature and coherence across proxy locations in (b)–(d) unforced and (e)–(g) forced model integrations. (a) Locations of all proxies (including marine sediment cores, tree-ring records, coral records, and ice cores) in the PAGES2k network (black). We indicate the sites of a North American tree-ring record (green triangle) and tropical Pacific Ocean coral $\delta O^{18}$ record (orange square). (b) Local surface temperature at the tree temperature-proxy site indicated by the green triangle in (a). Annual temperature anomalies are shown as green dots, with the 10-yr moving mean shown as a solid green line. Global mean surface-temperature anomalies (10-yr moving mean) shown in black. Temperature anomalies are calculated from modeled output from the CCSM4 piControl integration, which does not include any external forcings. To find local surface temperature, we interpolate surface-temperature output to the proxy location. Monthly temperature output is annually averaged and linearly detrended to calculate annual temperature anomalies. Standard deviation of annual mean local temperature is indicated in the top-right corner. (c) As in (b), but for the coral temperature-proxy site indicated by the orange square in (a). (d) Coherence between global temperature and local temperatures at the midlatitude tree temperature-proxy site in North America (green) and the coral temperature-proxy site in the tropical Pacific Ocean (orange). Spectra are calculated using Welch's method (Welch 1967) with 32 windows. Coherence at the 95% confidence level is calculated according to Amos and Koopmans (1963) and is indicated by the dashed black line. (e)–(g) As in (b)–(d), but using gridded temperature output from the CCSM last millennium integration, which includes external forcings.

Paleoclimate proxy records can reconstruct variability on time scales longer than the instrumental record (Huybers and Curry 2006; Ault et al. 2013; Laepple and Huybers 2014; Franzke et al. 2020), but care must be taken in interpreting them. In the sparse proxy record of the Common Era, for example (Fig. 1a), the degree to which a proxy time series is indicative of global climate change versus localized regional variations depends on the location of the proxy, the time scale of interest, and the underlying source of variability (e.g., Anchukaitis et al. 2012; Anchukaitis and Smerdon 2022). The relation between local and global temperature on different time scales can be measured by coherence (a frequency-dependent correlation metric; see section 3b). To illustrate this, we use paleoclimate simulations of the Community Climate

System Model, version 4 (CCSM4; Gent et al. 2011). We calculate coherence between global mean surface temperature and local surface temperature at two proxy sites in the Past Global Changes 2k (PAGES2k) database (PAGES2k Consortium 2017): that of a North American tree-ring record from San Francisco Peaks, and that of a tropical Pacific coral $\delta^{18}O$ record from Urvina Bay. Coherence is calculated using surface-temperature output from two CCSM4 simulations: an unforced preindustrial control run (Figs. 1b–d) and a forced last millennium run (Figs. 1e–g; Landrum et al. 2013).

We note that biology and joint dependencies on other climatic variables will further complicate how proxies at these two sites record temperature variability (Dee et al. 2017). All other things being equal, the results from this particular GCM suggest that low-latitude marine proxies such as corals are good recorders of unforced variability at interannual time scales but provide relatively little information about unforced variability on longer time scales (Fig. 1d, orange line). They do, however, record forced variability on these time scales (Fig. 1g, orange line). On the other hand, high-latitude continental records (e.g., trees) seem to primarily respond to regional variations on fast time scales, showing little coherence with global temperature on annual- to decadal-time scales (Fig. 1d, green line). Like low-latitude proxies, they are only able to provide information about forced global variations on long time scales (Fig. 1g, green line). Both types of proxies primarily reflect local temperature variability when no external forcing is present. However, both proxies have high coherence when external forcing is included, and thus largely reflect forced multidecadal global temperature variability rather than local processes.

Here, we seek to understand how forced and unforced variability are expected to differ in their manifestation across time and space using numerical climate models and basic physical principles. Such basic understanding is a necessary, though not sufficient, condition for both better interpreting paleoclimate variability and improving our mechanistic explanations of natural variability.

We fingerprint the spatial pattern of forced and unforced climate variability within CMIP5 (Taylor et al. 2012) preindustrial control and last millennium simulations (Braconnot et al. 2011), using two statistical metrics of annual mean surface temperature. First, we use frequency-dependent degrees of freedom ($\equiv N$; e.g., Kunz and Laepple 2021), which is related to the spatial covariance structure of temperature and broadly corresponds to the number of independently varying spatial patterns of temperature at a given time scale (section 3a). Second, we compute spectral coherence between local and global temperature (e.g., Mann and Park 1993; Chave et al. 1987; Lall and Mann 1995), which identifies regions that are driving global temperature variability at a given frequency (section 3b). Using these metrics, we characterize how the spatial patterns of temperature variability change as a function of frequency in forced and unforced simulations.

Care must be taken, however, in using GCM fingerprints. GCMs have been argued to reproduce global temperature reconstructions over the Common Era (Neukom et al. 2019; Zhu et al. 2019) and historical period (Fredriksen and Rypdal 2016), or even overestimate global mean temperature variability (Pallotta and Santer 2020) due to long-term trends in model-estimated internal variability (Po-Chedley et al. 2022). However, they do not match *regional* low-frequency variability, particularly in the low latitudes (Laepple and Huybers 2014; Bothe et al. 2015). To build confidence in the GCM results, we also explore the degree to which the salient features of the GCM fingerprint can be derived from simple energetic constraints. We develop a three-box Hasselmann-like model for temperature, with the flexibility to specify key climate parameters, including ocean heat capacity and climate-feedback strength (section 4a). We find that it qualitatively matches the CMIP5 results (section 4b): first, $N$ decreases at low frequencies. Our three-box model indicates that this response is due to temperature diffusion, as temperature gradients smooth out and temperature anomalies become more globally coherent on long time scales. Second, $N$ decreases at low frequencies when external forcing is included. This is because imposing global or near-global forcings, like volcanic eruptions, produces more globally orchestrated climatic responses. Decreases in $N$ correspond to increases in coherence. In section 4c, we also explore the impact of including more realistic, spatially varying ocean heat capacities and climate-feedback strengths in our three-box model. The results from our three-box model suggest that energetic constraints including downgradient energy transport may be a productive way to understand the time scale–dependent spatial structure of temperature variability. Conclusions are presented in section 5, where we also discuss potential applications of these findings to the proxy temperature record.

## 2. Multicentennial climate model integrations

We seek to evaluate forced and unforced climate variability as a function of frequency, and so we select nine fully coupled climate models from the CMIP5 project (Taylor et al. 2012) that each have both long preindustrial control (piControl) integrations (the shortest is 400 years), as well as last millennium (past1000) simulations. The nine models are BCC-CSM1.1, CCSM4, CSIRO Mk3L-1-2, FGOALS-s2, GISS-E2-R, IPSL-CM5A-LR, MIROC-ESM, MPI-ESM-P, and MRI-CGCM3. Monthly mean surface temperature output is annually averaged and regridded to a $128 \times 64$ grid (approximately $2.8^\circ$ resolution).

In piControl integrations, all external forcings (including from greenhouse gases, aerosols, ozone, and solar irradiance) are held constant at preindustrial levels. Thus, variability in these integrations is unforced, generated only through the internal climate dynamics represented in the models. These nine models also participated in the third Paleoclimate Model Intercomparison Project (PMIP3; Braconnot et al. 2011) and have past1000 simulations available. In the past1000 integrations, which are run from 850 to 1849 CE, orbital, solar, volcanic, greenhouse gas, and land-use forcings evolve through time, following Schmidt et al. (2011). Thus, variability in these integrations is a combination of forced and unforced variability, generated both through internal climate dynamics and climate responses to external (global) forcings.

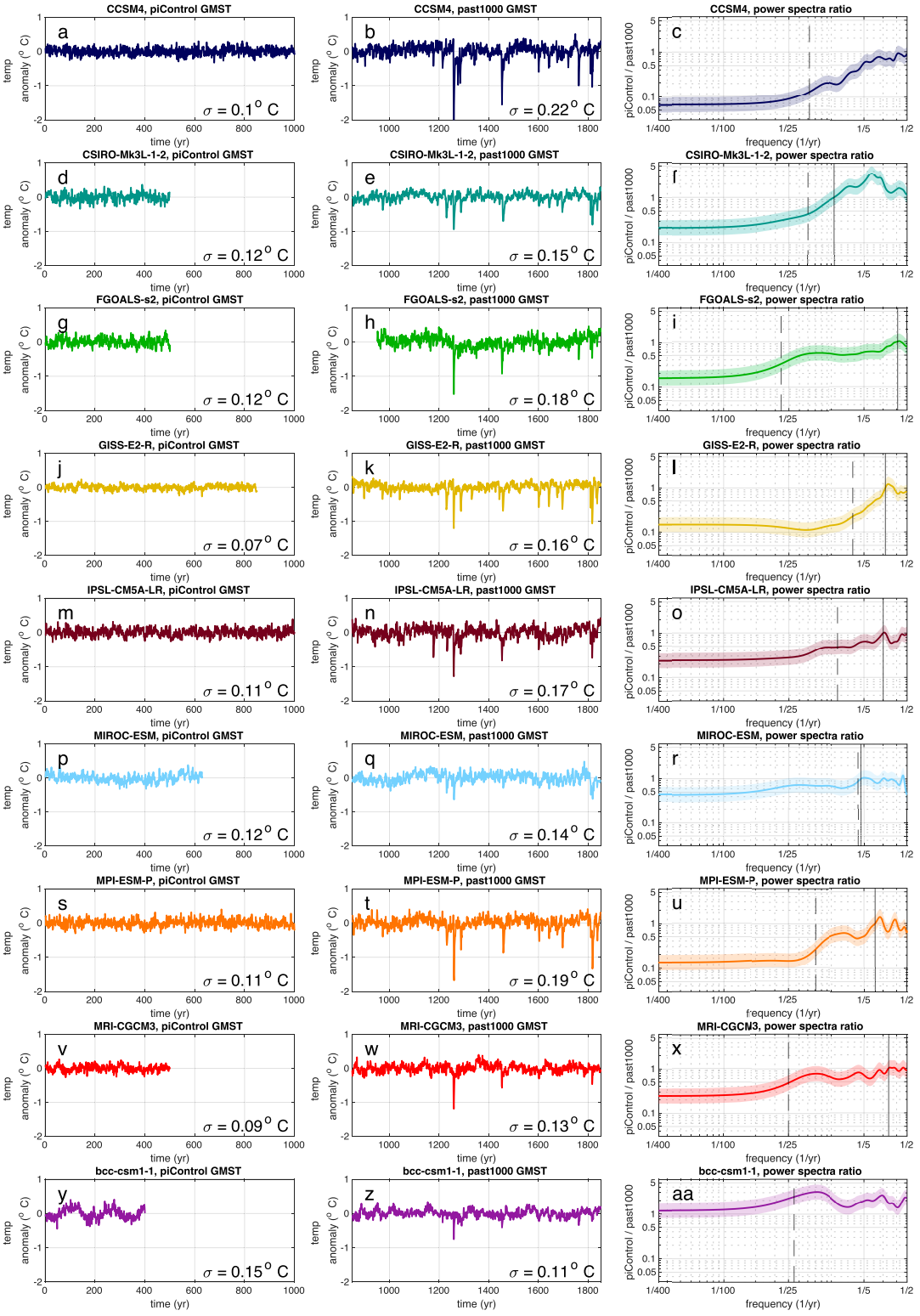


FIG. 2. GMST anomalies for nine CMIP5 (left) piControl and (center) past1000 simulations, and (right) relative contribution of unforced variability (piControl/past1000). The piControl simulations represent unforced variability, and the past1000 simulations represent forced variability driven primarily by volcanic eruptions. GMST is calculated as the area-weighted mean temperature. Monthly temperature data are annually averaged and linearly detrended to calculate annual temperature anomalies, shown in the first two columns. Black numbers in the bottom right corner indicate

In Fig. 2, we show the linearly detrended annual global mean surface temperature (GMST) anomalies for output from the piControl (Fig. 2, first column) and past1000 (Fig. 2, second column) integrations for all nine models. There is a wide range of internal GMST variability across the models; the standard deviation of annual mean GMST in the piControl simulations ranges between 0.07° and 0.15°C, with a mean value of 0.11°C. Notably, BCC-CSM1.1 has an apparent oscillation with a period of ~100 years, which is centered over Antarctica and the Southern Ocean (not shown). This oscillation disappears in the forced past1000 simulations. The spread in standard deviations across models is slightly larger in the past1000 simulations, ranging between 0.11° and 0.22°C, with an average value of 0.16°C. The strong negative temperature anomalies in the past1000 simulations correspond to volcanic eruptions, which are the primary mechanisms of forced variability in the simulations. Models with high overall variability in the past1000 simulations (e.g., CCSM4 in Fig. 2b; MPI-ESM-P in Fig. 2t) show strong volcanic responses, whereas others with lower overall variability (e.g., MIROC-ESM in Fig. 2q; BCC-CSM1.1 in Fig. 2z) have more muted volcanic responses. However, we note that intermodel comparison studies suggest that a model's volcanically forced temperature response primarily depends on model physics and chemistry (Clyne et al. 2021) rather than equilibrium climate sensitivity (Pauling et al. 2021).

To understand the role of external forcing on global variability at different time scales, we can compare the magnitude of externally forced variability to internal variability as a function of frequency. For each model, this approximation $\rho$ is computed as the ratio of the power spectrum of GMST in the piControl simulation to the power spectrum of GMST in the past1000 simulation (Fig. 2, third column). For all figures, power spectra are calculated using Welch's method and 32 windows. In testing with synthetic time series of red and pink noise, we found this provides a good smooth representation of the continuum spectrum without unduly affecting its shape.

Since past1000 simulations contain both forced and unforced variability, $\rho$ estimates the fractional contribution of unforced variability to total (forced and unforced) variability at each time scale. Across the CMIP5 models, unforced variability dominates on interannual time scales [frequencies $> (10\ \text{yr})^{-1}$], with $\rho$ approaching 1. Meanwhile, externally forced variability dominates at multidecadal time scales [frequencies $\leq (10\ \text{yr})^{-1}$], although there is a large spread among the models.

At the high end of this spread, in the BCC-CSM-1 model, $\rho = 1.2$ at low frequencies [$(400\ \text{yr})^{-1}$] (Fig. 2aa). This results from the strong multidecadal oscillations in the unforced simulations (Fig. 2y) that do not appear in the forced simulations (Fig. 2z). This is the only model in which internal variability dominates at all time scales ($\rho > 1$ at all frequencies), and the physical reasons for this behavior are unclear. At the low end of the spread, in the CCSM4 model, $\rho$ approaches 0.01 at low frequencies (Fig. 2c). This reflects the large sensitivity to volcanic forcing noted earlier. The ensemble-average low-frequency value of $\rho$ is 0.32. In some models, $\rho$ does not flatten out (dashed black lines in Fig. 2) until multidecadal time scales (e.g., CSIRO-Mk3L-1-2 in Fig. 2f; FGOALS-s2 in Fig. 2i), whereas in other models, it flattens out at higher frequencies (e.g., MIROC-ESM in Fig. 2r; GISS-E2-R in Fig. 2l). Models for which $\rho$ flattens out at lower frequencies are models in which adding external forcing adds more memory or redness relative to the unforced GMST power spectrum. The broad range of modeled behaviors in response to added external forcing suggests that different models have different mechanisms of low-frequency variability, different sensitivities to external forcing, or, as previously mentioned, different parameterizations of short-term volcanic forcing that projects onto low-frequency variability.

The time scale at which $\rho$ stops increasing (Fig. 2, dashed black line in third column) is qualitatively equivalent to the time scale at which the power spectra of past1000 simulations flattens. For all models included in this study, in both forced and unforced simulations, the power spectra of GMST flatten out by multidecadal to multicentennial time scales (Fig. S1 in the online supplemental material). There is, however, substantial evidence that both instrumental records (Fredriksen and Rypdal 2016; Franzke 2010) and paleoclimate proxies (e.g., Laepple and Huybers 2014; Nilsen et al. 2016; Hébert et al. 2022; Huybers and Curry 2006) exhibit a continuum of surface temperature variability that follows power-law scaling, or spectral power that continues to increase past multicentennial time scales. Some models with forcings on longer time scales (e.g., freshwater pulses; Zhu et al. 2019; Bakker et al. 2017) have produced power spectra that continue increasing past millennial time scales, so the failure of the CMIP5 models to exhibit power-law scaling may be because they do not include glacial variations.

## 3. Fingerprints of variability in CMIP5 models

### a. Degrees of freedom

We here characterize how the spatial patterns of variability change with frequency using a metric for the degrees of freedom. In statistics, the degrees of freedom $N_{\text{eff}}$ characterize the number of independent pieces of information in a dataset.

←
the standard deviation of GMST for that simulation. In the third column, the ratio of GMST piControl power spectrum divided by past1000 power spectrum yields an estimate of the relative contribution of unforced variability. Power spectra are estimated from the area-weighted, annual-mean GMST using Welch's method (Welch 1967) with 32 windows. Vertical solid black lines indicate the frequency at which external forcing begins to dominate over internal forcing (lowest frequency at which this ratio $> 1$). Vertical dashed black lines indicate the frequency at which this ratio begins to flatten out (ratio is twice its lowest-frequency value). Confidence at the 95% level (shaded) is calculated using a chi-squared approach. Colors correspond to different GCMs.

More informally, $N_{\text{eff}}$ can be thought of as the number of potential different modes by which the dynamical system might accommodate a perturbation.

For a time-varying climate field, $N_{\text{eff}}$ depends on the spatial correlation of the variations: a system with fewer degrees of freedom will have variability that is more globally coherent compared to a system with more regionally independent variability. We should also expect that $N_{\text{eff}}$ will depend on time scale: at longer time scales, we expect that different regions and components of the climate system will come into balance with each other and so will respond more coherently. Thus, $N_{\text{eff}}$ will be reduced at low frequencies. Previous work has shown this to be true in global climate models, observations (e.g., Jones et al. 1997), and simple energy balance models (North et al. 2011; Rypdal et al. 2015).

Here, we use a measure for $N_{\text{eff}}$ as a function of frequency, also suggested recently by Kunz and Laepple (2021). This metric is based on a ratio of regional to global variance density at each frequency. The variance of the mean of $N$ independent and identically distributed variables decreases with the number of variables:

$$\text{Var}(\langle T\rangle_N) = \frac{\text{Var}(T)}{N},$$

where $\text{Var}(T)$ is the variance of a random variable. Thus, a way to estimate the number of independent variations (i.e., the number of effective degrees of freedom) across time series at different spatial locations would be to take an estimate of the variance of each random variable $\langle \text{Var}(T)\rangle$ and divide it by the estimated variance of the average of all random variables $\text{Var}(\langle T\rangle_N)$,

$$N = \frac{\langle \text{Var}(T)\rangle}{\text{Var}(\langle T\rangle)N},$$

where the variance is taken across time and the average is taken across the number of spatial locations. However, this approach generates a single approximation of $N_{\text{eff}}$ across all frequencies.

To understand how $N_{\text{eff}}$ varies with time scale, we use power spectral densities (PSDs), which measure the variance contribution from a particular frequency. The ratio of regional to global PSD is a frequency-dependent ratio of regional to global variance. Let $P_{T_i}(f)$ be the PSD of the temperature at the $i$th grid point, and $P_{\langle T\rangle}(f)$ be the PSD of the area-weighted GMST $\langle T\rangle$. We then define $N_{\text{eff}}$ as the ratio of the mean of the local PSDs (area weighted) to the PSD of the global average temperature:

$$N_{\text{eff}}(f) = \frac{\langle P_{T_i}(f)\rangle}{P_{\langle T\rangle}(f)}, \tag{1}$$

which yields the mean variance of local temperature divided by the variance of GMST at each frequency. Hereafter, we drop the subscript "eff" after $N$. Uniform global temperature fluctuations would correspond to $N = 1$, while independent white noise in each grid cell would correspond to $N = 8921$ (our $128 \times 64$ grid).

Figure 3 shows $N(f)$ for both the unforced (Fig. 3a) and forced (Fig. 3b) simulations. Across CMIP5 models, $N$ decreases toward lower frequencies, indicating that variability becomes more globally coherent on longer time scales. At the highest frequency [$(2\text{ yr})^{-1}$], $N$ is similar in both the unforced and forced simulations [mean $N = 105$ (piControl); mean $N = 94$ (past1000)]. This suggests that on annual time scales, unforced variations dominate, even when external forcing is included. At the lowest frequency [$(400\text{ yr})^{-1}$], $N$ decreases by approximately fivefold in the unforced simulations (mean $N = 20$ for piControl) and by approximately 13-fold in the forced simulations (mean $N = 7$ for past1000). As noted above, this reduction in $N$ on long time scales is consistent with a diffusive system (Rypdal et al. 2015; Kunz and Laepple 2021), in which temperature anomalies smooth to larger spatial scales at lower frequencies.

Compared to the unforced simulations, the forced simulations consistently show reduced $N$ at low frequencies. Across models, at $(400\text{ yr})^{-1}$, $N$ is 3%–88% lower in the forced simulations, with an ensemble mean reduction of 61%. The smallest reduction in $N$ is in BCC-CSM1.1, which, as noted previously, has strong modes of internal variability that disappear in the forced simulations; all other models decrease by at least 36%. Lower $N$ in past1000 simulations indicates that slow variations in the forced simulations are dominated by external forcing that projects across large areas.

Overall, we see two primary features across the model ensemble: namely, reduced $N$ in forced simulations compared to the unforced and reduced $N$ at the low-frequency limit. This common behavior suggests that despite some model-to-model differences, relatively simple physical principles may govern how the structure of temperature variability changes with frequency.

### b. Coherence

The number of degrees of freedom are related to the number of independent sources of variability. To identify which regions and dynamical mechanisms drive global variability at different frequencies, we calculate the coherence of local surface temperature with GMST (e.g., Mann and Park 1993; Lall and Mann 1995; Chave et al. 1987). Coherence measures the fraction of variance at a given frequency shared between two time series. It is a frequency-dependent measure of cross correlation, bounded between zero and one. To ensure our coherence estimates are not unduly weighted toward the high latitudes, where variability is higher, we normalize GMST. Normalized GMST $\langle \tilde{T}\rangle = \sum_{i=1}^{8192}(A_i/A_{\text{Earth}})(T_i/\sigma_i)$, where $A_i/A_{\text{Earth}}$ is the fractional surface area, $T_i$ is the surface temperature, and $\sigma_i$ is the standard deviation of temperature in grid cell $i$. The normalization generates coherence estimates that identify regions that drive global variability (or are driven by global variability), rather than regions that have high-magnitude, but locally isolated, variability. We note that the coherence estimates using nonnormalized GMST are qualitatively similar to the coherence estimates shown in Fig. 4, although they are slightly higher in regions with low temperature variance (e.g., ~0.1 increase in the tropics zonal mean) and

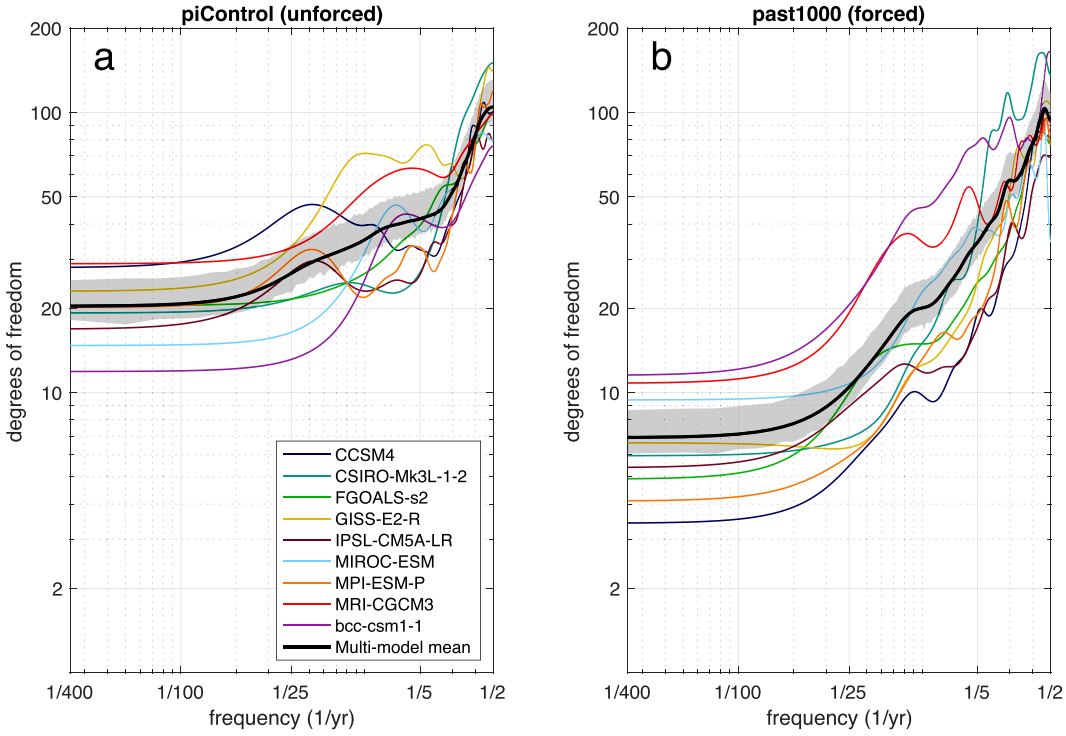


FIG. 3. Degrees of freedom $N$ in (a) piControl and(b) past1000 simulations. Degrees of freedom are frequency dependent, following Eq. (1). Spectral calculations are estimated using Welch's method with 32 windows. Colors correspond to different GCMs, with the multimodel mean shown in bold black. Shading indicates the mean of the models' 95% chi-squared confidence intervals.

slightly lower in regions with high temperature variance (e.g., ~0.15 decrease in the extratropics zonal mean).

Figure 4 presents the coherences calculated for both interdecadal [from $(2\ \text{yr})^{-1}$ to $(25\ \text{yr})^{-1}$] and multicentennial [from $(100\ \text{yr})^{-1}$ to $(400\ \text{yr})^{-1}$] frequency bands. We show the full spatial pattern in each of these frequency bands, along with zonal means. We expect that fewer instances of $N$ corresponding to higher mean coherence, and indeed, the mean coherence across CMIP5 models does demonstrate the same two primary features observed in the results for $N$ (i.e., Fig. 3). First, coherence increases at low frequencies: zonal mean coherence increases, on average, from 0.46 on interdecadal time scales (Fig. 4b) to 0.55 on multicentennial time scales (Fig. 4a) in the piControl simulations; and, in the past1000 simulations, from 0.47 on interdecadal time scales (Fig. 4d) to 0.72 on multicentennial time scales (Fig. 4c). This is consistent with a diffusive-like spatial smoothing of temperature fluctuations on long time scales. Second, coherence increases at the low-frequency limit when external forcing is present (cf. Fig. 4c and Fig. 5a). Zonal mean coherence is higher on multicentennial time scales in the past1000 simulations (Fig. 4c) than in the piControl simulations (Fig. 4a), except in the Southern Ocean (from approximately 70° to 40°S). This suggests that the Southern Ocean is locally driven, whereas global variability is driven by the tropics. When no external forcing is included, the Southern Ocean contributes heavily to global variability; when external forcing is included, however, the Southern Ocean accounts for a smaller proportion of global variability.

The spatial structure of coherence suggests that ENSO is a primary driver of global variability on interdecadal time scales in both forced and unforced simulations, as regions of high coherence (Figs. 4b,d) coincide with the ENSO pattern of sea surface temperature anomalies. Some echoes of the ENSO pattern remain at lower frequencies in the piControl simulations (Fig. 4a), so a reddened ENSO response may be driving internal variability on long time scales as well, as suggested by Newman et al. (2003). When external forcing is included, coherence on long time scales suggests that the tropics and midlatitudes (from 40°S to 40°N) are the primary drivers of global variability. There is particularly high coherence coinciding with the intertropical convergence zones (Fig. 4c), indicating that regions with deep convection (and, therefore, high precipitation) are coordinated with global temperature variability, consistent with previous results (Parsons et al. 2020).

Across all frequencies and forcing mechanisms, coherence with normalized GMST is highest in the tropics (Fig. 4). This suggests that temperature variability in the tropics is being efficiently communicated to the poles, whereas temperature variability in the high latitudes is not being communicated to the equator. The tropics' apparent control on global temperature variability is suggestive of downgradient (poleward) energy transport, which we explore in section 4.

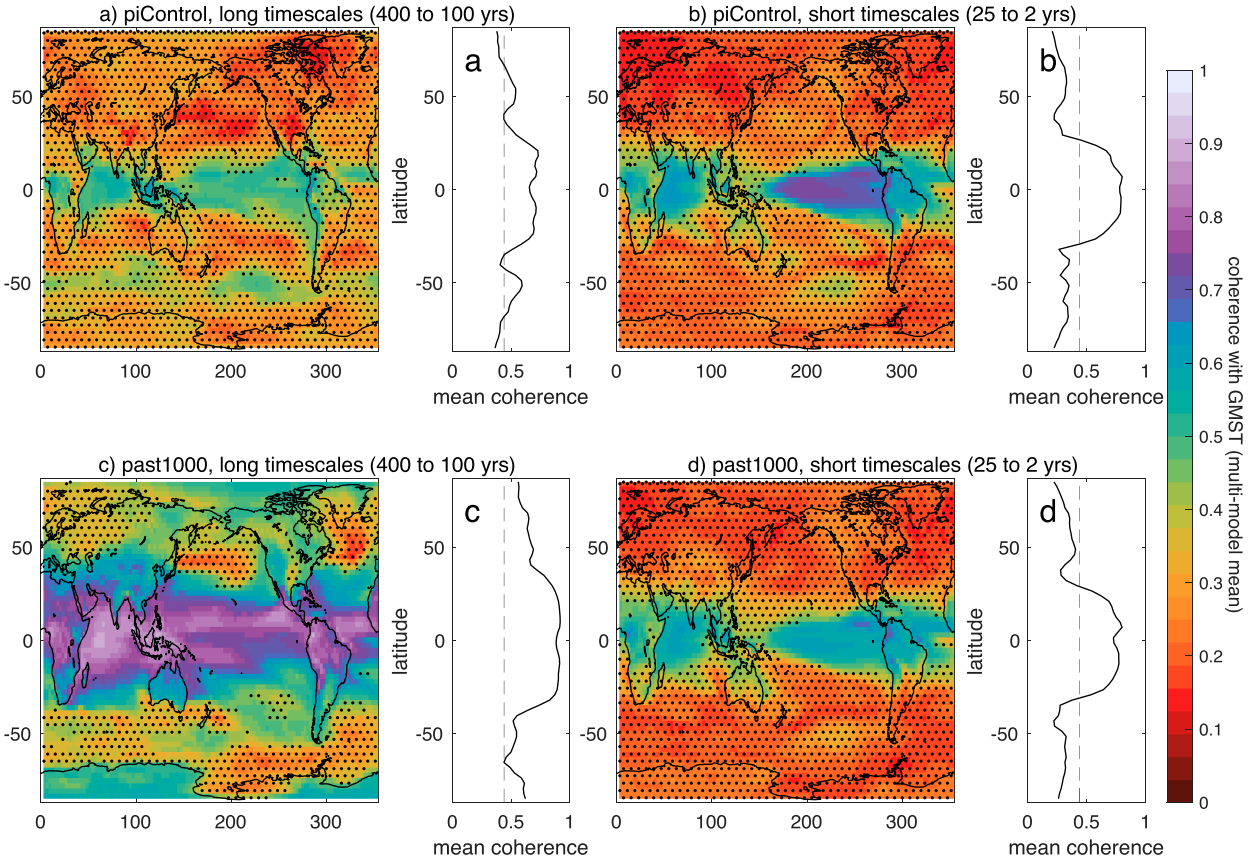


FIG. 4. Mean coherence between local temperature and normalized GMST in (a),(b) piControl simulations and (c),(d) past1000 simulations for (right) intra- to interdecadal variability [from $(2\ \mathrm{yr})^{-1}$ to $(25\ \mathrm{yr})^{-1}$] and (left) multicentennial variability [from $(100\ \mathrm{yr})^{-1}$ to $(400\ \mathrm{yr})^{-1}$] across CMIP5 models. Coherence between local temperature and normalized GMST (see text for details) is calculated for each individual model, and then coherence values are averaged across models. Stippling indicates grid cells that are insignificant at the 95% confidence level in at least five models. Coherence is calculated using Welch's method (Welch 1967) with 32 windows. Zonal-mean coherence (with coherence calculated between zonal mean temperature and GMST) is shown in panels to the right, with the 95% significance level indicated with a vertical dashed line.

## 4. Interpreting low-frequency dynamics with a conceptual model

In this section, we explore whether the behavior we have seen in this GCM ensemble can be understood as a result of constraints on Earth's energy budget. There are two key characteristics of climate variability across GCMs: 1) fewer degrees of freedom and greater spatial coherence at low frequencies than at high frequencies, especially when external forcing is present; and 2) at all time scales and in both forced and unforced simulations, tropical temperature variability is highly coherent with GMST variability. The first behavior is suggestive of diffusion-like dynamics (e.g., North et al. 2011; Rypdal et al. 2015), wherein a signal will spread over a larger area if given time. The second behavior is suggestive of a directionality in how temperature anomalies influence other regions. This can arise in the climate system due to the downgradient transport of moist static energy (MSE; sensible heat plus latent heat; e.g., Flannery 1984). The exponential increase of latent heat at warmer temperatures means that tropical temperature anomalies are associated with much greater energetic anomalies than high-latitude temperature anomalies of the same magnitude. These energetic anomalies are mixed downgradient and, as a result, anomalies in tropical temperatures are communicated nearly uniformly with latitude, whereas high-latitude anomalies remain much more regionally confined (Roe et al. 2015). Recent studies (e.g., Hwang and Frierson 2010; Rose et al. 2014; Roe et al. 2015; Siler et al. 2018) have shown that simple models governed by downgradient transport can largely characterize temperature, hydrology, and atmospheric heat transport in both the climatology and long-term climate response to anthropogenic climate forcing. The results of the previous two sections motivate exploring whether this downgradient energy-balance perspective may also help understand frequency-dependent behavior. To that end, we construct a simple three-box model based on principles of downgradient energy transport.

### a. *Three-box temperature model*

We create a simplified, stochastic, moist static energy balance model, based on the one-box Hasselmann (1976) model,

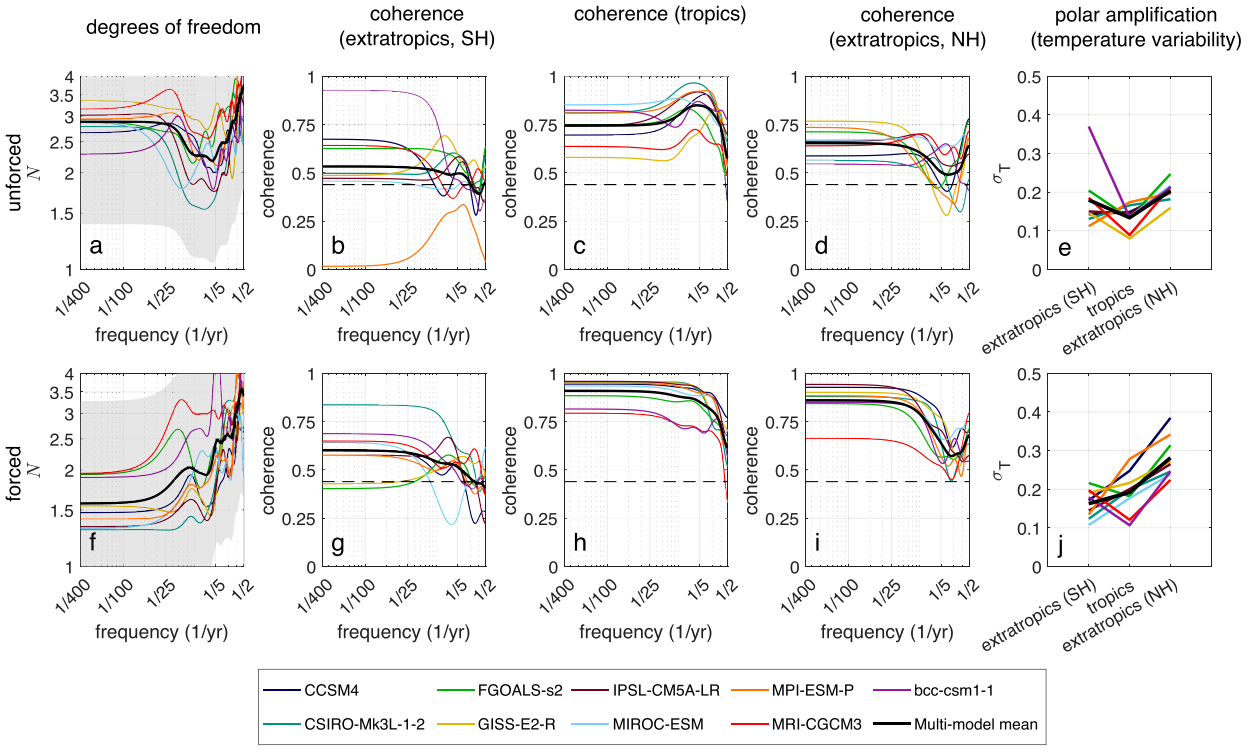


FIG. 5. Coarsened three-box CMIP5 results for (top) piControl simulations and (bottom) past1000 simulations. (a) Degrees of freedom, (b) coherence in the extratropical Southern Hemisphere, (c) coherence in the tropics, (d) coherence in the extratropical Northern Hemisphere, and (e) standard deviation of temperature in each box. Calculations are made after coarsening model output fields to three boxes (90°–30°S, 30°S–30°N, and 30°–90°N) in each piControl simulation. Coherence is calculated between local temperature and global mean surface temperature using Welch's method (Welch 1967) with 32 windows. Colors correspond to individual model results, and multimodel mean is shown in black. The chi-squared confidence interval for multimodel mean is shown at 95% level, representative of averaged individual model uncertainties. (f)–(j) As in (a)–(e), but for past1000 (forced) simulations.

adjusted to include moisture (Langen and Alexeev 2007) and extended to three boxes. This yields a modeled dynamical system that is simple enough to be solved analytically and allows the testing of parameter sensitivity, but retains the physical tendencies embodied in more complex, spatially resolved, moist static energy balance models (e.g., North et al. 1981).

Our model splits the world into three boxes: one tropical box (30°S–30°N) and two equal-sized extratropical boxes, one in the Southern Hemisphere (90°–30°S) and one in the Northern Hemisphere (30°–90°N). We consider temperature anomalies $T$ from some prescribed mean climate state. Each box is forced by stochastic radiative anomalies $R$ that may be globally coherent (e.g., due to volcanic eruptions $R_g$) or uncorrelated with other boxes (e.g., due to localized weather $R_{1,2,3}$). In each box, radiative anomalies can be dealt with in three ways: 1) radiated out to space; 2) exported to other boxes; or 3) integrated by the ocean. The importance of these pathways depends on the relative magnitude of radiative damping (governed by a climate-feedback parameter $\lambda$), the coupling between boxes (governed by a coupling parameter $\gamma$), and the ocean heat capacity $c$, which increases with ocean mixed layer depth. This yields the following set of equations:

$$c\frac{dT_1}{dt} = -\lambda T_1 + H_{21} + R_1 + R_g, \tag{3}$$

$$c\frac{dT_2}{dt} = -\lambda T_2 - H_{21} - H_{23} + R_2 + R_g, \quad \text{and} \tag{4}$$

$$c\frac{dT_3}{dt} = -\lambda T_3 + H_{23} + R_3 + R_g. \tag{5}$$

Here, $T_1$ and $T_3$ are the surface temperatures in the extratropics (box 1, Northern Hemisphere; box 3, Southern Hemisphere), and $T_2$ is the surface temperature in the tropics (box 2). Heat transfers between boxes are represented by $H_{21} = \gamma(\alpha T_2 - T_1)$ (transport from box 2 to box 1) and $H_{23} = \gamma(\alpha T_2 - T_3)$ (transport from box 2 to box 3) and depend on the temperature differential between the tropics and extratropics, the strength of $\gamma$, and the moisture constant $\alpha$, which allows for preferential transport of energy from the tropics to the extratropics when $\alpha > 1$ (Langen and Alexeev 2007). The moisture constant is based on the ratio of the MSE increase in the tropics to the extratropics when temperature warms 1°C, and is approximately 1.7, based on current climatology. We define the resulting standard deviation of temperature in each box as $\sigma_{T_i}$ $(i = 1, 2, 3)$.

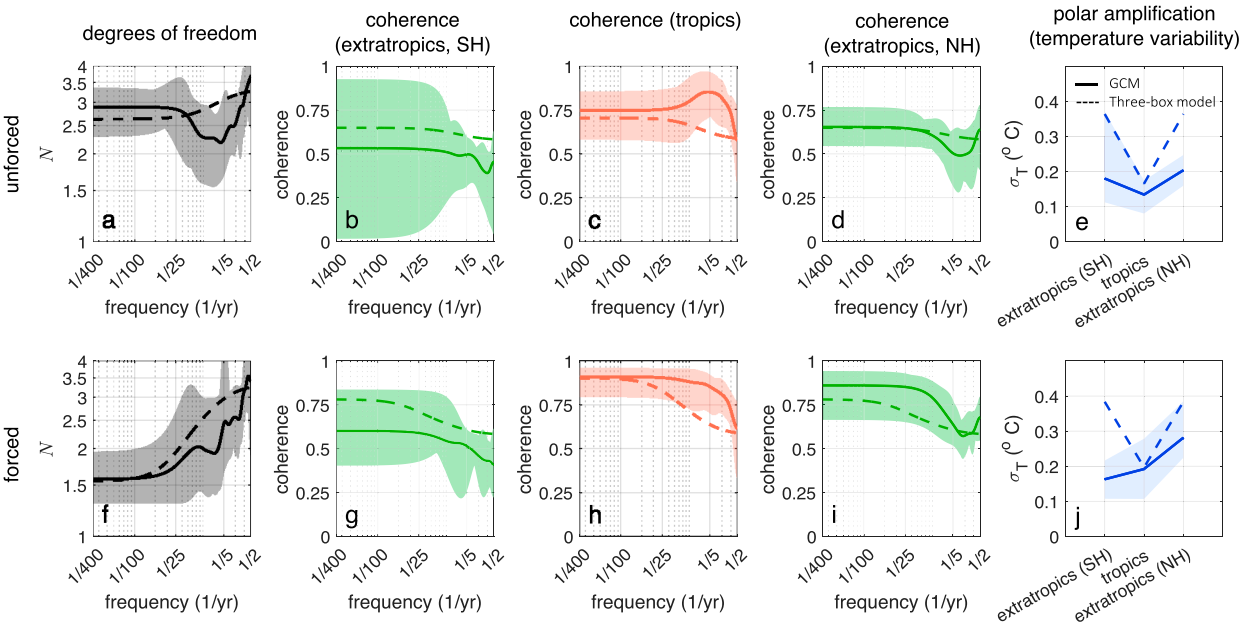


FIG. 6. Coarsened three-box CMIP5 results for (top) piControl simulations and (bottom) past1000 simulations, and three-box model results without global forcing in the top panels and with global forcing in the bottom panels. (a) Degrees of freedom, (b) coherence in the extratropical Southern Hemisphere, (c) coherence in the tropics, (d) coherence in the extratropical Northern Hemisphere, and (e) standard deviation of temperature in each box. Results are shown both for the coarsened three-box CMIP5 results (Fig. 5; solid lines indicate multimodel mean, and shading indicates full model spread), and for the three-box model results (dashed lines). For the three-box model, parameter values are chosen to be physically realistic and match high- and low-frequency behavior in coarsened GCMs ($c = 2$ W m$^{-2}$ K$^{-1}$ yr$^{-1}$, $\gamma = 0.15$ W m$^{-2}$ K$^{-1}$, $\lambda = 1.5$ W m$^{-2}$ K$^{-1}$, $\alpha = 1.7$, $\sigma_{R_{1,3}} = 1$ W m$^{-2}$, $\sigma_{R_2} = 0.5$ W m$^{-2}$, $\sigma_{R_g} = 0$ (unforced), or $\sigma_{R_g} = 0.2$ (forced) W m$^{-2}$). Global forcing is applied as red noise with a lag-1 correlation coefficient $r = 0.2$. For CMIP5 results, coherence is calculated between local temperature and global-mean surface temperature using Welch's method (Welch 1967) with 32 windows. (f)–(j) As in (a)–(e), but for past1000 (forced) simulations.

We can normalize Eqs. (3)–(5) to find the following four key parameters that govern the behavior of our three-box model:

1) $\gamma/\lambda$, which is a measure of the relative efficiency of horizontal heat transport to radiative damping;
2) $\alpha$, which controls how much more efficiently heat is transported out of the tropics to the extratropics, associated with the latent heat;
3) the characteristic time scale $c/\lambda$, which controls the frequency at which the power spectra of $T_{1,2,3}$ and GMST flatten; and
4) $\kappa = \sigma^2_{R_{1,2,3}}/(\sigma^2_{R_{1,2,3}} + \sigma^2_{R_g})$, the relative importance of local radiative anomalies to total radiative anomalies.

Equations (3)–(5) can be solved analytically for the degrees of freedom $N$ and the coherence (see the appendix). In the following sections, we explore the dynamics of the three-box model, including its frequency-dependent behavior ($N$ and coherence) and the standard deviation of temperature $\sigma_{T_i}$.

### b. Fitting conceptual model to CMIP5 output

We want to evaluate how the three-box system compares to the variability of the GCM. To do this, in each of the nine CMIP5 models, we average (i.e., coarsen) the GCM output fields over the same three areas included in our three-box model: the Southern Hemisphere extratropics, the tropics, and the Northern Hemisphere extratropics. In effect, we are analyzing GCM variability at the hemispheric scale. Figure 5 shows the resulting $N$ (Figs. 5a,f), coherence (Figs. 5b–d,g–i), and $\sigma_{T_{1,2,3}}$ (Figs. 5e,j) in the coarsened GCM.

The coarsened CMIP5 models exhibit $N$ with qualitatively similar behavior to the full CMIP5 models. $N$ decreases toward low frequencies [mean $N$ decreases from $N = 3.69$ at $(2\text{ yr})^{-1}$ to $N = 2.88$ at $(400\text{ yr})^{-1}$ in the piControl simulations (Fig. 5a) and from $N = 3.41$ at $(2\text{ yr})^{-1}$ to $N = 1.57$ at $(400\text{ yr})^{-1}$ in the past1000 simulations (Fig. 6f)], especially when external forcing is present (cf. Fig. 5a and Fig. 5f). We note that $N$ has an upper bound of 4 because our model treats the tropics as two (coherent) regions and there are two independent extratropical regions (see appendix for analytical solutions). Finally, an equator-to-pole gradient in $\sigma_{T_i}$ is present (Figs. 5e,j), which steepens when external forcing is present (cf. Fig. 5e and Fig. 6j).

We also note some other features of the coarsened GCM results. Across all piControl simulations, $N$ decreases in the frequency band from $\sim(5\text{ yr})^{-1}$ to $(7\text{ yr})^{-1}$ (Fig. 6a), which can be explained by coherent temperature anomalies associated with ENSO. Coherence between tropical temperature and GMST is maximized within this same frequency band, indicating that tropical temperature anomalies are driving variability at this frequency, which is consistent with an ENSO sea surface

temperature pattern. Coherence between extratropical temperature and GMST is minimized in the Northern Hemisphere in the ENSO frequency band (Fig. 5d), while the Southern Hemisphere is largely unaffected by ENSO (Fig. 5b).

There are also significant hemispheric differences in the extratropical response to external forcing. On average, coherence in the Southern Hemisphere is relatively unchanged when external forcing is present (solid black line in Fig. 5b, cf. with Fig. 5g), while coherence in the Northern Hemisphere increases [primarily at frequencies of not more than $\sim(10\ \mathrm{yr})^{-1}$] and temperature variability increases in almost all models when external forcing is present (cf. Fig. 5d and Fig. 5i; cf. Fig. 5e and Fig. 5j). This suggests that variability in the Southern Hemisphere is largely governed by internal dynamics, as also indicated in Fig. 4. However, we note that there are large intermodel spreads in Southern Hemisphere coherence, especially in the piControl runs (Fig. 5b), which indicates that different models have significantly different representations of internal dynamics in the Southern Hemisphere.

To evaluate whether our three-box model can capture GCM behavior, we fit it to the coarsened GCM results. Our parameters are chosen to be physically realistic and match high- and low-frequency behavior in the coarsened GCMs. The parameter set is $\lambda = 1.5\ \mathrm{W\,m^{-2}\,K^{-1}}$, equivalent to a global mean equilibrium climate sensitivity of 2.7 K for a 4 $\mathrm{W\ m^{-2}}$ forcing; $\gamma = 0.15\ \mathrm{W\,m^{-2}\,K^{-1}}$; $\alpha = 1.7$; $c = 2\ \mathrm{W\,m^{-2}\,K^{-1}}$ yr, equivalent to a mixed layer depth of 15 m; and, for the forced runs, local forcing that accounts for 80% of total extratropical forcing and 60% of total tropical forcing. Additional parameter details are included in the caption for Fig. 6. Our chosen parameters yield $\gamma/\lambda = 0.1$, while past studies constraining dry diffusivity have used $\gamma/\lambda$ values closer to 1 (e.g., Armour et al. 2019; Langen and Alexeev 2007). The characteristic time scale ($c/\lambda$) implied by these parameters is 1.3 yr.

Figure 6 shows solutions for $N$ and coherence in the three-box model as a function of frequency (Fig. 6, dashed lines) and also $\sigma_{T_{1,2,3}}$ (Figs. 6e,j), which measures polar amplification of temperature variability. For our parameters, $N$ decreases from 3.3 at the high-frequency limit to 2.6 at the low-frequency limit (unforced; Fig. 6a), and from 3.3 at the high-frequency limit to 1.6 at the low-frequency limit (forced; Fig. 6f), plateauing to minimum values at approximately $(25\ \mathrm{yr})^{-1}$ in both simulations. This is consistent with diffusion acting to distribute temperature anomalies more uniformly on longer time scales. Similarly, in all three boxes, coherence with GMST increases at low frequencies (Figs. 6b–d,g–i) as temperature anomalies become more spatially uniform.

At all frequencies, the tropical box ($T_2$; Figs. 6c,h) has higher coherence with the global mean ($\overline{T}$) than either of extratropical boxes ($T_{1,3}$; Figs. 6b,d,g,i). The tropical box is half the surface area of the globe, and temperature anomalies are coherent across the tropics. By contrast, each extratropical box is only a quarter of the global surface area, and $R_1$ and $R_3$ are independent.

Both $N$ and coherence show good agreement between coarsened GCMs and the three-box model, in both forced and unforced simulations (Figs. 6a–d,f–i). Agreement is especially good at the high- and low-frequencies, suggesting that frequency-dependent variability in GCMs can be characterized by a diffusive process. A relatively small amount of global forcing ($\sigma_{R_g} = 0.2\ \mathrm{W\,m^{-2}}$, accounting for 20% of total forcing in the extratropics and 40% of total forcing in the tropics) leads to a dramatic reduction in low-frequency $N$ (cf. Fig. 6a and Fig. 7f). Our three-box model is not able to capture GCM behavior on decadal and multidecadal time scales [from $\sim(5\ \mathrm{yr})^{-1}$ to $(25\ \mathrm{yr})^{-1}$], which is indicative of mechanisms not included in our model (i.e., ENSO). We also note that sources of local variability (e.g., lateral variations in temperature) are averaged out in the coarsened GCM, which limits interpretation of the parameter values used in the three-box model.

We also note that this standard set of parameters produces polar-amplified temperature variability ($\sigma_{T_{1,3}} > \sigma_{T_2}$). This is produced through two mechanisms. First, it is easier for heat anomalies to spread from the tropics to the extratropics than from one extratropical region to the other extratropical region in the opposite hemisphere, because the tropics can exchange heat with both extratropical boxes while each extratropical box can only exchange heat with the tropics. The extratropics are geographically isolated, compared to the tropics. Second, the MSE gradient $\alpha$ allows for a preferential transport of energy from the tropics to the extratropics, which amplifies temperature anomalies in the extratropics. We note that polar amplification typically refers to the phenomenon where, under external radiative forcing, the magnitude of long-term surface temperature change at high latitudes is higher than the GMST change. Here, we point toward the existence of polar-amplified temperature variability, which does not require external forcing. Moreover, our model does not include spatially varying feedback patterns or ocean mixed layer depths (i.e., $\lambda$ and $c$ are fixed constants) but is able, nonetheless, to produce amplified temperature responses in the extratropics on long time scales. This suggests that polar amplification of temperature variability is at least partially due to 1) the relative geographic isolation of the extratropics (i.e., the extratropics are only connected to the tropics, whereas the tropics are connected to both extratropical regions), and 2) downgradient energy transport.

In both the three-box model and the coarsened GCMs, Northern Hemisphere polar amplification is enhanced when external forcing is added (cf. Fig. 6e and Fig. 7j). However, polar amplification is suppressed in the Southern Hemisphere in the coarsened, forced GCM simulations (Fig. 6j), which the three-box model is unable to emulate. The hemispheric differences in coherence and temperature variability suggest that spatially varying $\lambda$ and $c$ affect GCM behavior.

### c. Sensitivity analysis

While the coarsened GCM and three-box model share the same fundamental, diffusive-like behavior, there remains a large spread among GCM estimates of frequency-dependent behavior. Next, we explore how $N$, coherence, and $\sigma_{T_{1,2,3}}$ vary as a function of $\gamma/\lambda$, $\alpha$, $\gamma/\lambda$, and $\kappa$ to understand the extent to which these physical parameters might explain the intermodel spread. Readers not interested in the effect of varying parameters

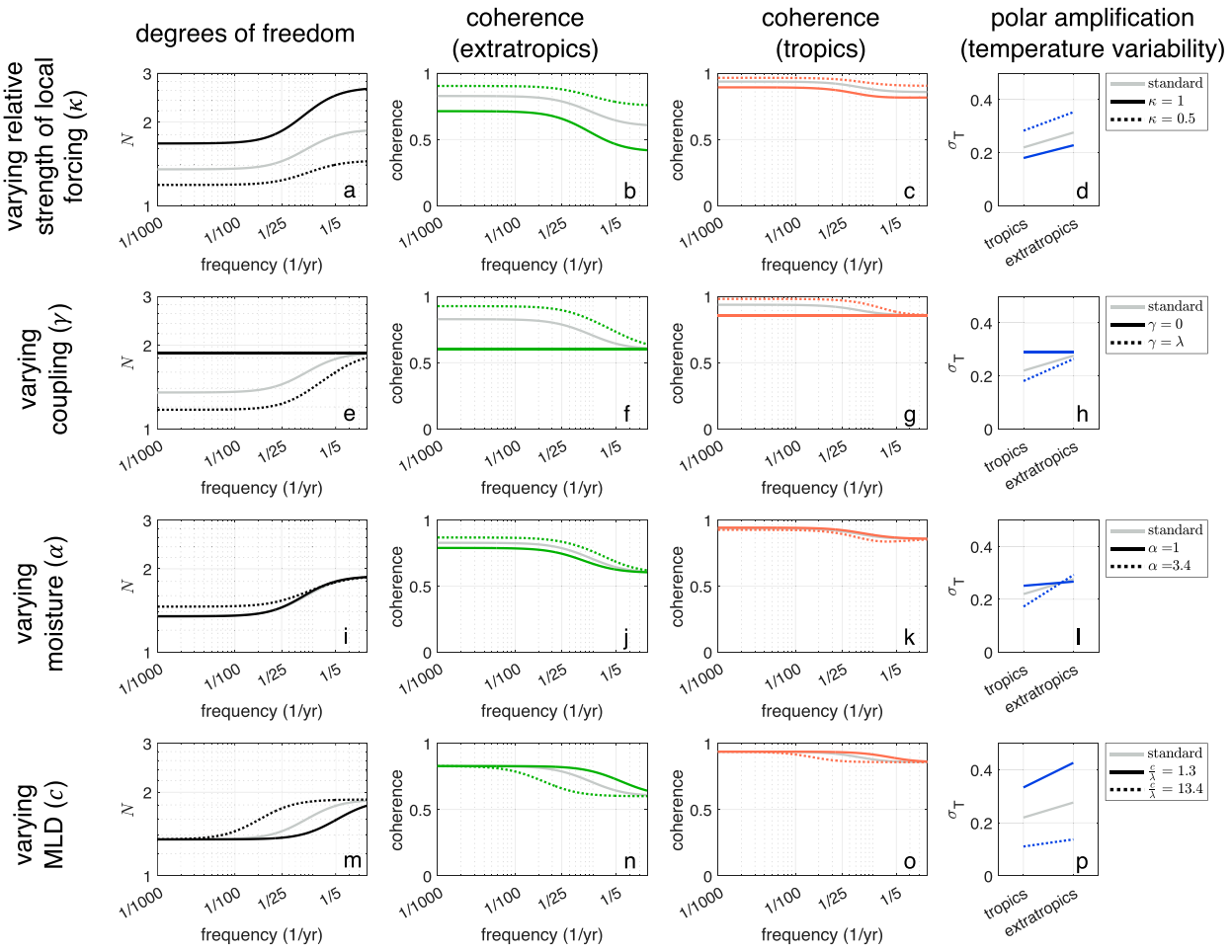


FIG. 7. Degrees of freedom, spatially varying coherence, and polar amplification in the three-box model as a function of parameter set. Impact of varying $\kappa$ (i.e., relative strength of global forcing) from $\kappa = 1$ (solid) to $\kappa = 0.5$ (dashed) on (a) degrees of freedom, (b) coherence in extratropics, (c) coherence in tropics, and (d) standard deviation of temperature in tropics and extratropics, which is a measure of polar amplification of variability. Solid gray lines are for comparison and show results for a standard parameter set with values: $\lambda = 1.5$ W m$^{-2}$ K$^{-1}$; $\gamma = 0.5$ W m$^{-2}$ K$^{-1}$; $\alpha = 1.7$; $c = 5$ W m$^{-2}$ K$^{-1}$ yr$^{-1}$, equivalent to a mixed layer depth of 40 m; and $\kappa = 0.75$, or local forcing that accounts for 75% of total forcing. (e)–(h) As in (a)–(d), but for varying coupling strength from $\gamma = 0$ to $\gamma = \lambda$. (i)–(l) As in (a)–(d), but for varying moisture parameter from $\alpha = 1$ to 3.4. (m)–(p) As in (a)–(d), but for varying mixed layer depth from $c/\lambda = 1.3$ to 13.4 yr$^{-1}$, which is equivalent to varying mixed layer depths from 15 to 150 m. These curves are plotted from the analytic solutions to the three-box model that are provided in the appendix.

on the three-box model can skip to section 5. For this section, we probe the parameters away from a simplified standard set (detailed in the Fig. 7 caption), which is distinct from the parameter set used to fit to coarsened GCM results shown in Fig. 6.

As $\kappa$ decreases, $N$ decreases and coherence increases at all frequencies (Figs. 7a–c) because temperature responses have become more globally orchestrated. Since the temperature anomalies between boxes are more coherent with each other, heat transfer between boxes is suppressed, and the equator-to-pole gradient in temperature variability flattens (Fig. 7d).

As $\gamma/\lambda$ increases, $N$ decreases and coherence increases on long time scales (Figs. 7e–g). Temperature anomalies in individual boxes are being less efficiently transported to other boxes, which results in less coherent temperature anomalies between boxes. When there is no coupling ($\gamma/\lambda = 0$), $N$ and coherence are constant across all frequencies, since there is no heat exchange between boxes on any time scale. Polar amplification increases with $\gamma/\lambda$ (Fig. 7h), because the mechanisms for amplified temperature responses in the extratropics identified in the previous section (preferential transport of energy from the tropics to the extratropics when $\alpha > 1$, and the trapping of heat at the extratropics due to the relative geographic isolation of the extratropics) can only operate when heat transfer between boxes is allowed.

As the MSE gradient steepens ($\alpha$ increases) and poleward transport enhances relative to equatorward transport, $N$ increases (Fig. 7i). This suggests that energy becomes trapped in the extratropics, which are incoherent with each other, driving an increase in $N$ primarily on longer time scales. Due to the influx of heat from the tropics, the extratropical boxes

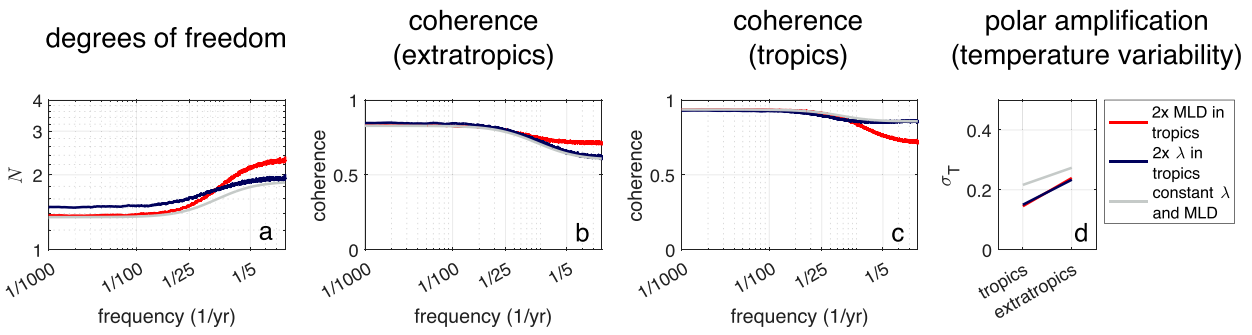


FIG. 8. Degrees of freedom, spatially varying coherence, and polar amplification in the three-box model with spatially varying feedbacks and mixed layer depth. (a) Degrees of freedom, (b) coherence in the extratropics, (c) coherence in the tropics, and (d) standard deviation of temperature in the tropics and extratropics. In one experiment (red), $c$ is twice as large in the tropics relative to the extratropics. In the other (blue), $\lambda$ is twice as large in the tropics relative to the extratropics. Results for the standard parameter set are shown in gray for comparison. Parameter values for the standard parameter set are described in Fig. 7. For all cases, the area-weighted mean $c$ and $\lambda$ are the same as in the standard parameter set. Coherence is estimated using Welch's method (Welch 1967) with 32 windows.

contribute more heavily to GMST relative to the dry model, leading to enhanced coherence in the extratropics (Fig. 7j) and slightly lower coherence in the tropics (Fig. 7k). Similarly, polar amplification is magnified when moisture is present (Fig. 7l). As $\alpha$ increases, the same export of heat from the tropics requires a smaller tropical temperature fluctuation. The relative polar amplification is thus predominantly a suppression of tropical temperature variability, rather than a boosting of extratropical variability.

As the characteristic time scale $c/\lambda$ decreases, the frequency at which $N$ and coherence plateau shifts to higher values (shorter time scales; Figs. 7m–o) because temperature anomalies are more quickly integrated by the ocean. High-frequency temperature variability in all three boxes is also enhanced, resulting in added polar amplification as temperature anomalies are more quickly trapped at the poles (Fig. 7p).

Finally, we briefly explore the impact of spatially varying $\lambda$ and $c$ on $N$, coherence, and $\sigma_{T_i}$ (Fig. 8). We run two experiments: 1) setting tropical ocean heat capacity to be twice as large as extratropical ocean heat capacity ($c_2 = 2c_{1,3}$; Fig. 8, red lines), which mimics the spatial pattern of summertime ocean heat capacity (e.g., Chang et al. 2013); and 2) setting the tropical radiative feedback to be twice as large (and, therefore, more stable) as the extratropical radiative feedback ($\lambda_2 = 2\lambda_{1,3}$; Fig. 8, blue lines), which mimics the effect of adding a sea ice–albedo feedback in the extratropics.

When $c_2 = 2c_{1,3}$, $N$ increases at high frequencies but has the same low-frequency behavior as in the standard parameter set (Fig. 8a). High-frequency extratropical coherence is higher, and high-frequency tropical coherence is lower, relative to the standard parameter set (Figs. 8b,c), because temperature anomalies are more slowly integrated by an ocean with a deeper mixed layer depth (higher $c$). Polar amplification increases (Fig. 8d; cf. gray line and red line) because high-frequency variability in the extratropics is enhanced and high-frequency variability in the tropics is suppressed. Similarly, when $\lambda_2 = 2\lambda_{1,3}$, polar amplification increases (Fig. 8d; cf. gray line and blue line) because temperature anomalies in the extratropics are not as effectively damped compared with the standard parameter set. Spatially varying $\lambda$ and $c$ results in changes to $N$, coherence, and polar amplification that are comparable to changes produced by varying the mean values of parameters within uncertainty bounds.

## 5. Discussion and summary

We have evaluated frequency-dependent behavior across a suite of long-running CMIP5 GCMs simulations from two experiments, one of which only contains internal variability and the other of which is subject to external forcing [the preindustrial control (piControl) experiment and last millennium (past1000) experiment, respectively]. Specifically, across a range of frequencies, we evaluated $N$, which estimates the number of independently varying spatial patterns of temperature anomalies, and coherence with global mean temperature, which identifies regions driving variability. There are distinct fingerprints of variability in forced and unforced simulations, which may be useful in understanding how temperature variability is expressed in the paleoclimate record and evaluating how well temperature variability in GCMs maps onto temperature variability observed in paleoclimate proxy data.

We identified three features robust across GCMs: 1) decreased $N$ at low frequencies; 2) an additional reduction in low-frequency $N$ when external forcing is included; and 3) a high coherence with global mean surface temperature in the tropics. The first two features are indicative of temperature diffusion or signals spreading over larger areas over longer time periods. The third feature is indicative of downgradient energy transport, wherein temperature anomalies are efficiently communicated from the tropics to the extratropics due to sensible and latent heat fluxes. While these tendencies are generally robust across models, there are large model-to-model differences in their intensity and associated time scales (e.g., Figs. 2, 3, and 5). In other words, the mechanisms and intensity of low-frequency variability continue to be poorly constrained among GCM ensembles, although, as previously mentioned, some of the intermodel spread in forced simulations may be due to differences in how high-frequency volcanic forcing is applied and projected onto longer time scales (e.g., Clyne et al. 2021). We further note that our characterizations of

forced temperature variability are derived from the past1000 simulations, whereby forcing is primarily generated through volcanic eruptions, and our findings may change for a different forcing and background climate state. A further step could be evaluating spatial fingerprints in response to specific forcings using single-forcing ensembles.

We showed that the frequency-dependent behavior in GCMs is broadly consistent with a three-box moist static energy balance model. Our model is a function of four key climate parameters: ocean heat capacity, diffusivity, the strength of radiative damping, and the relative importance of local forcing to total forcing. Understanding how different parameters change the shape of $N$ and coherence provides insight into the mechanisms controlling intermodel differences in frequency-dependent behavior. Our model also points toward differing behavior in the northern and southern extratropics; specifically, in GCMs, temperature variability in the Southern Hemisphere is relatively independent of ENSO and appears to be largely governed by local internal dynamics.

We also evaluated the spatial pattern of temperature variability in our three-box model and in the GCMs. We identified two mechanisms for polar amplification of temperature variability: 1) the bidirectional heat transport from the tropics and the relative geographic isolation of the extratropics (resulting in unidirectional heat transport from the extratropics); and 2) the moist static energy gradient. These allow for the preferential transportation of energy poleward and the trapping of additional energy in the extratropics, which enhances extratropical temperature variability relative to the tropics. Temperature power spectra are thus redder in the extratropics and whiter in the tropics.

Our characterizations of temperature variability in forced and unforced simulations suggest that diffusive tendencies may help interpret the proxy paleoclimate record over the last millennium. Varying physical parameters within the three-box model has a large impact on frequency-dependent behavior and provides possible explanations for model-to-model differences in $N$ and coherence. Understanding which models have physics most consistent with proxy paleoclimate reconstructions is a key next step. However, we note that proxies encode climate data imperfectly, which further alters signals of variability. Proxies are sensitive to nonclimate processes, including ice-core down-core diffusion (Dee et al. 2017), measurement noise, and bioturbation (Laepple and Huybers 2013). Proxies also may have seasonal biases or joint dependencies (e.g., some tree-ring width records may reflect the influence of both temperature and precipitation; Tolwinski-Ward et al. 2011; Fritts 1966; St. George and Ault 2014) or systematically biased spatial coverage (e.g., Judd et al. 2020), all of which influence recorded variability. In addition, proxies may record highly localized variability that cannot be resolved by GCMs (e.g., Laepple and Huybers 2014). Thus, interpreting proxy reconstructions using our characterizations of temperature variability ultimately must consider these complicating factors, including how the proxy network itself will filter those signals of variability. Here, we primarily focus on the dynamics driving temperature variability, although understanding how our characterizations of forced and unforced variability are expressed in the proxy network will ultimately require explicit consideration of proxy noise and filtering processes.

Nevertheless, our work suggests some clear targets for distinguishing between forced and unforced temperature variability, which may be reflected in the paleoclimate proxy record. In particular, redder (steeper) spectral slopes and enhanced spatial coherence are hallmarks of forced variability. Relatedly, models suggest that some locations (e.g., the tropics) may be more indicative of global temperature variability and others (e.g., the extratropics) are more indicative of regional variability. Our analyses also show that these spatial coherences are a function of forcing and time scale, knowledge of which may help improve covariances used in data assimilation algorithms. We propose that much of this behavior can be understood in terms of diffusive, downgradient energy transport, which is consistent with the existence of polar-amplified, low-frequency temperature variability. Finally, although it has not been a focus of this paper, that same downgradient transport has been shown to orchestrate spatial patterns of both temperature and hydrology (e.g., Siler et al. 2018). Hydrology and polar-amplified temperature variability could, in principle, provide additional metrics for evaluating proxy records.

*Acknowledgments.* The authors acknowledge support from NSF P2C2 2102829 (C.P. and G.H.R.), NSF AGS 2019647 (G.H.R.), University of Washington Program on Climate Change Graduate Student Fellowship (R.C.S.), the U.S. Science Support Program and Schlanger Fellowship (R.C.S.), and the University of Washington Provost Fellowship (R.C.S.). We thank Dennis Hartmann, Kevin Anchukaitis, and Thomas Laepple and his research group for their comments.

*Data availability statement.* All CMIP5 output can be downloaded from the Earth System Grid Federation (https://esgf-node.llnl.gov/search/cmip5/). The three-box moist static energy balance model is available at https://github.com/beccacs/fingerprinting_variability. Other analyses are available from the authors upon request.

## APPENDIX

### Analytic Solutions for the Three-Box Model

#### *a. Three-box model*

The equations of the three-box model are

$$c\frac{dT_1}{dt} = -\lambda T_1 + H_{21} + R_1 + R_g, \tag{A1}$$

$$c\frac{dT_2}{dt} = -\lambda T_2 - H_{21} - H_{23} + R_2 + R_g, \tag{A2}$$

$$c\frac{dT_3}{dt} = -\lambda T_3 + H_{23} + R_3 + R_g, \tag{A3}$$

$$H_{21} = \gamma(\alpha T_2 - T_1), \quad \text{and} \tag{A4}$$

$$H_{23} = \gamma(\alpha T_2 - T_3), \tag{A5}$$

where $T_1$, $T_2$, and $T_3$ are surface temperatures in box 1 (Northern Hemisphere extratropics), box 2 (tropics), and box 3

(Southern Hemisphere extratropics), respectively; $\lambda$ is a climate feedback parameter, $c$ is the ocean heat capacity, and $R_i$ is radiative anomaly with standard deviation $\sigma_{R_i}$ ($i = 1, 2, 3, g$). Hereafter, the frequency and forcings are normalized by $\lambda$ such that $\omega' = (c/\lambda)\omega$ and $F_i' = F_i/\lambda$ for $i = 1, 2, 3, g$, where $\omega = 2\pi f$, such that Eqs. (A1)–(A5) are transformed to have units of temperature (K). The prime symbols are dropped in subsequent expressions. Setting $R_g = 0$ approximates the (unforced) piControl simulations, whereas setting $R_g \neq 0$ approximates the (forced) past1000 simulations.

Equations (A1)–(A5) can be solved for various useful metrics, including degrees of freedom, coherence, and the magnitude and frequency dependence of polar amplified temperature variability.

### b. Degrees of freedom

To compute degrees of freedom [Eq. (1)], we need expressions for the power spectrum of temperature in each box ($P_{T_1}$, $P_{T_2}$, $P_{T_3}$), and an expression for the power spectrum of area-weighted global mean temperature $\langle P_T \rangle = P_{(T_1+2T_2+T_3)/4}$, where the tropical box accounts for one-half of Earth's surface area and $\langle \cdot \rangle$ denotes an average or expected value. The power spectra $P_{T_i}$ can also be written as an autospectrum $\phi_{T_iT_i}$,

$$P_{T_i}(\omega) = \phi_{T_iT_i}(\omega) = \langle \tilde{T}_i(\omega)\tilde{T}_i^*(\omega) \rangle, \tag{A6}$$

where a tilde indicates a Fourier transform and an asterisk denotes a complex conjugate. For completeness we provide the full solutions, which are algebraically messy. In appendix section c, we simplify by considering a symmetric case and some limiting cases of high and low frequency. We assume the forcing terms $R_1$, $R_2$, $R_3$, and $R_g$ are independent white-noise forcing, such that $\phi_{R_iR_j} = 0$ when $i \neq j$, and $\phi_{R_iR_i} = \sigma_{R_i}^2$. The solutions below can be easily modified to include independent red-noise forcing by substituting $\sigma_i = \sqrt{2R_i^2/\{r[1 + (\omega^2/r^2)]\}}$, where the lag-1 autocorrelation between two successive time samples has correlation coefficient $0 < r < 1$.

The power spectrum of mean global temperature can be written as

$$\begin{aligned}\phi_{\langle T\rangle\langle T\rangle}(\omega) &= P_{\langle T\rangle}(\omega) \\ &= (\sigma_1^2 + \sigma_3^2)\frac{\{\omega^2 + [1 + 2(1+\alpha)\gamma]^2\}}{16(\omega^2+1)[\omega^2 + (1+\gamma+2\alpha\gamma)^2]} \\ &\quad + \sigma_2^2\frac{\omega^2 + (1+\gamma+\alpha\gamma)^2}{4(\omega^2+1)[\omega^2+(1+\gamma+2\alpha\gamma)^2]} \\ &\quad + \sigma_g^2\frac{4\omega^2 + [2+3(1+\alpha)\gamma]^2}{4(\omega^2+1)[\omega^2+(1+\gamma+2\alpha\gamma)^2]}.\end{aligned} \tag{A7}$$

Power spectra of individual boxes can be written as

$$\begin{aligned}\phi_{T_1T_1}(\omega) &= P_{T_1}(\omega) \\ &= \frac{1}{\omega^2+1}\left\{\sigma_{R_1}^2\frac{(\alpha\gamma^2+2\alpha\gamma+\gamma+1)^2 + \omega^4 + \omega^2[(4\alpha^2+2\alpha+1)\gamma^2 + (4\alpha+2)\gamma+2]}{[\omega^2+(\gamma+1)^2][\omega^2+(2\alpha\gamma+\gamma+1)^2]}\right. \\ &\quad \left. + \sigma_{R_2}^2\frac{(\alpha\gamma)^2}{[\omega^2+(1+\gamma+2\alpha\gamma)^2]} + \sigma_{R_3}^2\frac{\alpha^2\gamma^4}{[\omega^2+(\gamma+1)^2][\omega^2+(1+\gamma+2\alpha\gamma)^2]} + \sigma_{R_g}^2\frac{[\omega^2+(1+3\alpha\gamma)^2]}{[\omega^2+(1+\gamma+2\alpha\gamma)^2]}\right\},\end{aligned} \tag{A8}$$

$$\phi_{T_2T_2}(\omega) = P_{T_2}(\omega) = \frac{(\sigma_{R_1}^2+\sigma_{R_3}^2)\gamma^2 + \sigma_{R_2}^2[(\gamma+1)^2+\omega^2] + \sigma_{R_g}^2[(3\gamma+1)^2+\omega^2]}{(\omega^2+1)[\omega^2+(1+\gamma+2\alpha\gamma)^2]}, \tag{A9}$$

$$\phi_{T_3T_3}(\omega) = P_{T_3}(\omega) = \frac{1}{\omega^2+1}\left\{\begin{aligned}&\sigma_{R_1}^2\frac{\alpha^2\gamma^4}{[\omega^2+(\gamma+1)^2][\omega^2+(1+\gamma+2\alpha\gamma)^2]} + \sigma_{R_2}^2\frac{(\alpha\gamma)^2}{[\omega^2+(1+\gamma+2\alpha\gamma)^2]} \\ &+ \sigma_{R_3}^2\frac{(\alpha\gamma^2+2\alpha\gamma+\gamma+1)^2+\omega^4+\omega^2[(4\alpha^2+2\alpha+1)\gamma^2+(4\alpha+2)\gamma+2]}{[\omega^2+(\gamma+1)^2][\omega^2+(2\alpha\gamma+\gamma+1)^2]} \\ &+ \sigma_{R_g}^2\frac{[\omega^2+(1+3\alpha\gamma)^2]}{[\omega^2+(1+\gamma+2\alpha\gamma)^2]}\end{aligned}\right\}. \tag{A10}$$

The degrees of freedom $N$ are equal to $\langle P_T \rangle / P_{\langle T \rangle}$, or $(P_{T_1} + 2P_{T_2} + P_{T_3})/4P_{\langle T \rangle}$, following Eq. (1).

### c. Coherence

To compute coherence between global-mean temperature and temperature in box $i$, an estimate of the cross-power density $\phi_{T_i\langle T\rangle}$ between the two is needed:

$$\phi_{T_i\langle T\rangle}(\omega) = \langle \tilde{T}_i(\omega)\tilde{T}^*(\omega) \rangle, \tag{A11}$$

or, for each individual box:

$$\phi_{T_1\langle T\rangle}(\omega) = \sigma_{R_1}^2 \frac{[1+2(1+\alpha)\gamma + i\omega][\alpha\gamma^2 + (1+2\alpha)\gamma(1-i\omega) - (i+\omega)^2]}{4(1+\gamma-i\omega)(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_2}^2 \frac{\alpha\gamma(1+\gamma+\alpha\gamma+i\omega)}{2(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_3}^2 \frac{\alpha\gamma^2[1+2(1+\alpha)\gamma+i\omega]}{4(1+\gamma-i\omega)(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_g}^2 \frac{(1+3\alpha\gamma-i\omega)(2+3(1+\alpha)\gamma+2i\omega)}{2(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)}, \tag{A12}$$

$$\phi_{T_2\langle T\rangle}(\omega) = \sigma_{R_1}^2 \frac{\gamma[1+2(1+\alpha)\gamma+i\omega]}{4(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_2}^2 \frac{(1+\gamma-i\omega)(1+\gamma+\alpha\gamma+i\omega)}{2(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_3}^2 \frac{\gamma[1+2(1+\alpha)\gamma+i\omega]}{4(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_g}^2 \frac{(1+3\gamma-i\omega)[2+3(1+\alpha)\gamma+2i\omega]}{2(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)}, \tag{A13}$$

$$\phi_{T_3\langle T\rangle}(\omega) = \sigma_{R_1}^2 \frac{\alpha\gamma^2[1+2(1+\alpha)\gamma+i\omega]}{4(1+\gamma-i\omega)(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_2}^2 \frac{\alpha\gamma(1+\gamma+\alpha\gamma+i\omega)}{2(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_3}^2 \frac{[1+2(1+\alpha)\gamma+i\omega][\alpha\gamma^2 + (1+2\alpha)\gamma(1-i\omega) - (i+\omega)^2]}{4(1+\gamma-i\omega)(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)} + \sigma_{R_g}^2 \frac{(1+3\alpha\gamma-i\omega)[2+3(1+\alpha)\gamma+2i\omega]}{2(1+\gamma+2\alpha\gamma-i\omega)(1+\gamma+2\alpha\gamma+i\omega)(1+\omega^2)}. \tag{A14}$$

The coherence can be written as a complex function,

$$C_{T_i\langle T\rangle}(\omega) = \frac{\phi_{T_i\langle T\rangle}(\omega)}{\sqrt{\phi_{T_iT_i}(\omega)\phi_{\langle T\rangle\langle T\rangle}(\omega)}} \tag{A15}$$

where $\phi_{T_iT_i}$ and $\phi_{\langle T\rangle\langle T\rangle}$ come from Eqs. (A8)–(A10) and (A7), respectively.

*d. Example: Symmetrical world and limiting cases*

In the simplest case of a symmetric world ($\sigma_{R_1} = \sigma_{R_2} = \sigma_{R_3}$), then the degrees of freedom can be written as

$$\frac{P_{\langle T\rangle}}{\langle P_T\rangle}(\omega) = \frac{8[(4+\alpha+3\alpha^2)\gamma^3(-3+2\kappa) + 3(1+\alpha^2)\gamma^4(-3+2\kappa)]}{[\omega^2+(1+\gamma)^2]\{2\gamma(2\kappa-3)(1+\alpha)[4+3(1+\alpha)\gamma+(-8+5\kappa)(1+\omega^2)]\}} + \frac{8\{2\gamma[-5+\alpha(-3+\kappa)+2\kappa](1+\omega^2) - 2(1+\omega^2)^2\}}{[\omega^2+(1+\gamma)^2]\{2\gamma(2\kappa-3)(1+\alpha)[4+3(1+\alpha)\gamma+(-8+5\kappa)(1+\omega^2)]\}} + \frac{8(\gamma^2\{-23-11\omega^2+\alpha^2(-9+4\kappa)(1+\omega^2)+2\kappa(7+3\omega^2)-2\alpha[6+\kappa(-3+\omega^2)]\})}{[\omega^2+(1+\gamma)^2]\{2\gamma(2\kappa-3)(1+\alpha)[4+3(1+\alpha)\gamma+(-8+5\kappa)(1+\omega^2)]\}} \tag{A16}$$

where $\kappa$ is the relative contribution of local forcing to total forcing, $\sigma_{R_i}^2/(\sigma_g^2 + \sigma_{R_i}^2)$ for $i = 1, 2, 3$.

To understand how energy diffuses throughout the model over time, we can examine the high- and low-frequency limits of $N$ ($N_{\omega\to\infty}$ and $N_{\omega\to 0}$). The limit $N_{\omega\to\infty}$ is a function only of $\kappa$, with higher $\kappa$ associated with higher $N_{\omega\to\infty}$, because, on short time scales, temperature anomalies cannot be exported out of the box. When local forcing in the extratropics dominates ($\sigma_{R_{1,3}}^2 \gg \sigma_{R_2}^2$), $N_{\omega\to\infty} = 4$. When local forcing in the tropics dominates ($\sigma_{R_2}^2 \gg \sigma_{R_{1,3}}^2$), $N_{\omega\to\infty} = 2$. When global forcing dominates ($\sigma_{R_g}^2 \gg \sigma_{R_{1,2,3}}^2$), $N_{\omega\to\infty} = 1$. The exact functional dependence of the low-frequency limit $N_{\omega\to 0}$ is more complicated but depends on $\kappa$, $\gamma/\lambda$, and $\alpha$, with lower $N_{\omega\to 0}$ under parameter combinations that favor efficient heat transport.

Equation (A16) can be additionally simplified if we assume a dry atmosphere ($\alpha = 1$):

$$\frac{P_{\langle T\rangle}}{\langle P_T\rangle}(\omega) = \frac{8(3\gamma^4 + 8\gamma^3)(2\kappa - 3) + 2\gamma^2[(2\kappa - 5)\omega^2 + 6\kappa - 11]}{[\omega^2 + (1+\gamma)^2][8(2\gamma + 3\gamma^2)(2\kappa - 3) + (5\kappa - 8)(\omega^2 + 1)]} + \frac{8[\gamma(3\kappa - 8)(\omega^2 + 1) - (\omega^2 + 1)^2]}{[\omega^2 + (1+\gamma)^2][8(2\gamma + 3\gamma^2)(2\kappa - 3) + (5\kappa - 8)(\omega^2 + 1)]}. \quad \text{(A17)}$$

As $\gamma \to 0$, $P_{\langle T\rangle}/\langle P_T\rangle \to 8/(8 - 5\kappa)$, which is 1 when $\sigma_1^2 = 0$ and 8/3 when $\sigma_g^2 = 0$. As $\gamma \to \infty$, $P_{\langle T\rangle}/\langle P_T\rangle \to 1$.

At high frequencies ($\omega \to \infty$), $P_{\langle T\rangle}/\langle P_T\rangle \to 8/(8 - 5\kappa)$. At low frequencies ($\omega \to 0$),

$$\frac{P_{\langle T\rangle}}{\langle P_T\rangle} \to \frac{8[-1 + 8\gamma^3(-3 + 2\kappa) + \gamma(-8 + 3\kappa) + 2\gamma^2(-11 + 6\kappa) + \gamma^4(-9 + 6\kappa)]}{(1+\gamma)^2[-8 + 5\kappa + 16\gamma(2\kappa - 3) + 24\gamma^2(2\kappa - 3)]}$$

As $\kappa \to 1$ (such that $\sigma_{R_1}^2 \gg \sigma_{R_g}^2$),

$$\frac{P_{\langle T\rangle}}{\langle P_T\rangle} \to \frac{8[8\gamma^3 + 3\gamma^4 + 5\gamma(\omega^2 + 1) + (\omega^2 + 1)^2 + 2\gamma^2(5 + 3\omega^2)]}{[\omega^2 + (1+\gamma)^2](3 + 16\gamma + 24\gamma^2 + 3\omega^2)}$$

As $\kappa \to 0$ (such that $\sigma_{R_1}^2 \ll \sigma_{R_g}^2$), $P_{\langle T\rangle}/\langle P_T\rangle \to 1$.

Even when $\alpha = 1$, there is some polar amplification. The ratio of the power spectra of temperature in the extratropics $P_{T_1}$ to that in the tropics $P_{T_2}$ is

$$\frac{P_{T_1}}{P_{T_2}}(\omega) = \frac{(2\kappa - 3)(8\gamma^3 + 3\gamma^4) + 2\gamma(\kappa - 4)(1 + \omega^2) - (1 + \omega^2)^2 + 2\gamma^2[-11 - 5\omega^2 + \kappa(5 + \omega^2)]}{[-1 + 2\gamma(2\kappa - 3)(1 + \gamma) - \omega^2][\omega^2 + (1+\gamma)^2]}. \quad \text{(A18)}$$

## REFERENCES


Amos, D. E., and L. H. Koopmans, 1963: *Tables of the Distribution of the Coefficient of Coherence for Stationary Bivariate Gaussian Processes*. Sandia Corporation, 328 pp.

Anchukaitis, K. J., and J. E. Smerdon, 2022: Progress and uncertainties in global and hemispheric temperature reconstructions of the Common Era. *Quat. Sci. Rev.*, **286**, 107537, https://doi.org/10.1016/j.quascirev.2022.107537.

——, and Coauthors, 2012: Tree rings and volcanic cooling. *Nat. Geosci.*, **5**, 836–837, https://doi.org/10.1038/ngeo1645.

Armour, K. C., N. Siler, A. Donohoe, and G. H. Roe, 2019: Meridional atmospheric heat transport constrained by energetics and mediated by large-scale diffusion. *J. Climate*, **32**, 3655–3680, https://doi.org/10.1175/JCLI-D-18-0563.1.

Ault, T. R., J. E. Cole, J. T. Overpeck, G. T. Pederson, S. S. George, B. Otto-Bliesner, C. A. Woodhouse, and C. Deser, 2013: The continuum of hydroclimate variability in western North America during the last millennium. *J. Climate*, **26**, 5863–5878, https://doi.org/10.1175/JCLI-D-11-00732.1.

Bakker, P., P. U. Clark, N. R. Golledge, A. Schmittner, and M. E. Weber, 2017: Centennial-scale Holocene climate variations amplified by Antarctic ice sheet discharge. *Nature*, **541**, 72–76, https://doi.org/10.1038/nature20582.

——, H. Goosse, and D. M. Roche, 2022: Internal climate variability and spatial temperature correlations during the past 2000 years. *Climate Past*, **18**, 2523–2544, https://doi.org/10.5194/cp-18-2523-2022.

Bothe, O., and Coauthors, 2015: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Climate Past*, **11**, 1673–1699, https://doi.org/10.5194/cp-11-1673-2015.

Braconnot, P., S. P. Harrison, B. L. Otto-Bliesner, A. Abe-Ouchi, J. Jungclaus, and J.-Y. Peterschmitt, 2011: The Paleoclimate Modeling Intercomparison Project contribution to CMIP5. *CLIVAR Exchanges*, No. 16, International CLIVAR Project Office, Southampton, United Kingdom, 15–19.

Brown, P. T., W. Li, and S.-P. Xie, 2015: Regions of significant influence on unforced global mean surface air temperature variability in climate models. *J. Geophys. Res. Atmos.*, **120**, 480–494, https://doi.org/10.1002/2014JD022576.

Chang, Y. S., S. Zhang, A. Rosati, T. L. Delworth, and W. F. Stern, 2013: An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. *Climate Dyn.*, **40**, 775–803, https://doi.org/10.1007/s00382-012-1412-2.

Chave, A. D., D. J. Thomson, and M. E. Ander, 1987: On the robust estimation of power spectra, coherences, and transfer functions. *J. Geophys. Res.*, **92**, 633–648, https://doi.org/10.1029/JB092iB01p00633.

Clyne, M., and Coauthors, 2021: Model physics and chemistry causing intermodel disagreement within the VolMIP-Tambora interactive stratospheric aerosol ensemble. *Atmos. Chem. Phys.*, **21**, 3317–3343, https://doi.org/10.5194/acp-21-3317-2021.

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

DelSole, T., 2006: Low-frequency variations of surface temperature in observations and simulations. *J. Climate*, **19**, 4487–4507, https://doi.org/10.1175/JCLI3879.1.

Flannery, B. P., 1984: Energy balance models incorporating transport of thermal and latent energy. *J. Atmos. Sci.*, **41**, 414–421, https://doi.org/10.1175/1520-0469(1984)041<0414:EBMITO>2.0.CO;2.

Franzke, C. L. E., 2010: Long-range dependence and climate noise characteristics of Antarctic temperature data. *J. Climate*, **23**, 6074–6081, https://doi.org/10.1175/2010JCLI3654.1.

——, and Coauthors, 2020: The structure of climate variability across scales. *Rev. Geophys.*, **58**, e2019RG000657, https://doi.org/10.1029/2019RG000657.

Fredriksen, H. B., and K. Rypdal, 2016: Spectral characteristics of instrumental and climate model surface temperatures. *J. Climate*, **29**, 1253–1268, https://doi.org/10.1175/JCLI-D-15-0457.1.

Fritts, H. C., 1966: Growth-rings of trees: Their correlation with climate: Patterns of ring widths in trees in semiarid sites depend on climate-controlled physiological factors. *Science*, **154**, 973–979, https://doi.org/10.1126/science.154.3752.973.

Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1.

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

Hausfather, Z., H. F. Drake, T. Abbott, and G. A. Schmidt, 2020: Evaluating the performance of past climate model projections. *Geophys. Res. Lett.*, **47**, e2019GL085378, https://doi.org/10.1029/2019GL085378.

Hébert, R., U. Herzschuh, and T. Laepple, 2022: Millennial-scale climate variability over land overprinted by ocean temperature fluctuations. *Nat. Geosci.*, **15**, 899–905, https://doi.org/10.1038/s41561-022-01056-4.

Hegerl, G. C., and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 663–745.

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

Hwang, Y. T., and D. M. W. Frierson, 2010: Increasing atmospheric poleward energy transport with global warming. *Geophys. Res. Lett.*, **37**, L24807, https://doi.org/10.1029/2010GL045440.

Jones, P. D., T. J. Osborn, and K. R. Briffa, 1997: Estimating sampling errors in large-scale temperature averages. *J. Climate*, **10**, 2548–2568, https://doi.org/10.1175/1520-0442(1997)010<2548:ESEILS>2.0.CO;2.

Judd, E. J., T. Bhattacharya, and L. C. Ivany, 2020: A dynamical framework for interpreting ancient sea surface temperatures. *Geophys. Res. Lett.*, **47**, e2020GL089044, https://doi.org/10.1029/2020GL089044.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1.

Kunz, T., and T. Laepple, 2021: Frequency-dependent estimation of effective spatial degrees of freedom. *J. Climate*, **34**, 7373–7388, https://doi.org/10.1175/JCLI-D-20-0228.1.

Laepple, T., and P. Huybers, 2013: Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability. *Earth Planet. Sci. Lett.*, **375**, 418–429, https://doi.org/10.1016/j.epsl.2013.06.006.

——, and ——, 2014: Ocean surface temperature variability: Large model-data differences at decadal and longer periods. *Proc. Natl. Acad. Sci. USA*, **111**, 16 682–16 687, https://doi.org/10.1073/pnas.1412077111.

Lall, U., and M. Mann, 1995: The Great Salt Lake: A barometer of low-frequency climatic variability. *Water Resour. Res.*, **31**, 2503–2515, https://doi.org/10.1029/95WR01950.

Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom, and H. Teng, 2013: Last millennium climate and its variability in CCSM4. *J. Climate*, **26**, 1085–1111, https://doi.org/10.1175/JCLI-D-11-00326.1.

Langen, P. L., and V. A. Alexeev, 2007: Polar amplification as a preferred response in an idealized aquaplanet GCM. *Climate Dyn.*, **29**, 305–317, https://doi.org/10.1007/s00382-006-0221-x.

Mann, M., and J. Park, 1993: Spatial correlations of interdecadal variation in global surface temperatures. *Geophys. Res. Lett.*, **20**, 1055–1058, https://doi.org/10.1029/93GL00752.

Neukom, R., and Coauthors, 2019: Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nat. Geosci.*, **12**, 643–649, https://doi.org/10.1038/s41561-019-0400-0.

Newman, M., G. P. Compo, and M. A. Alexander, 2003: ENSO-forced variability of the Pacific decadal oscillation *J. Climate*, **16**, 3853–3857, https://doi.org/10.1175/1520-0442(2003)016<3853:EVOTPD>2.0.CO;2.

Nilsen, T., K. Rypdal, and H. Fredriksen, 2016: Are there multiple scaling regimes in Holocene temperature records? *Earth Syst. Dyn.*, **7**, 419–439, https://doi.org/10.5194/esd-7-419-2016.

North, G. R., R. F. Cahalan, and J. A. Coakley, 1981: Energy balance climate models. *Rev. Geophys.*, **19**, 91–121, https://doi.org/10.1029/RG019i001p00091.

——, J. Wang, and M. G. Genton, 2011: Correlation models for temperature fields. *J. Climate*, **24**, 5850–5862, https://doi.org/10.1175/2011JCLI4199.1.

PAGES2k Consortium, 2017: A global multiproxy database for temperature reconstructions of the Common Era. *Sci. Data*, **4**, 170088, https://doi.org/10.1038/sdata.2017.88.

Pallotta, G., and B. D. Santer, 2020: Multi-frequency analysis of simulated versus observed variability in tropospheric temperature. *J. Climate*, **33**, 10 383–10 402, https://doi.org/10.1175/JCLI-D-20-0023.1.

Parsons, L. A., M. K. Brennan, R. C. J. Wills, and C. Proistosescu, 2020: Magnitudes and spatial patterns of interdecadal temperature variability in CMIP6. *Geophys. Res. Lett.*, **47**, e2019GL086588, https://doi.org/10.1029/2019GL086588.

Pauling, A. G., M. Bushuk, and C. M. Bitz, 2021: Robust interhemispheric asymmetry in the response to symmetric volcanic forcing in model large ensembles. *Geophys. Res. Lett.*, **48**, e2021GL092558, https://doi.org/10.1029/2021GL092558.

Po-Chedley, S., J. T. Fasullo, N. Siler, Z. M. Labe, E. A. Barnes, C. J. W. Bonfils, and B. D. Santer, 2022: Internal variability and forcing influence model–satellite differences in the rate of tropical tropospheric warming. *Proc. Natl. Acad. Sci. USA*, **119**, e2209431119, https://doi.org/10.1073/pnas.2209431119.

Roe, G. H., N. Feldl, K. C. Armour, Y. T. Hwang, and D. M. W. Frierson, 2015: The remote impacts of climate feedbacks on regional climate predictability. *Nat. Geosci.*, **8**, 135–139, https://doi.org/10.1038/ngeo2346.

Rose, B. E. J., K. C. Armour, D. S. Battisti, N. Feldl, and D. D. B. Koll, 2014: The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophys. Res. Lett.*, **41**, 1071–1078, https://doi.org/10.1002/2013GL058955.

Rypdal, K., M. Rypdal, and H. B. Fredriksen, 2015: Spatiotemporal long-range persistence in Earth's temperature field: Analysis of stochastic-diffusive energy balance models. *J. Climate*, **28**, 8379–8395, https://doi.org/10.1175/JCLI-D-15-0183.1.

Schmidt, G. A., and Coauthors, 2011: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geosci. Model Dev.*, **4**, 33–45, https://doi.org/10.5194/gmd-4-33-2011.

Siler, N., G. H. Roe, and K. C. Armour, 2018: Insights into the zonal-mean response of the hydrologic cycle to global warming from a diffusive energy balance model. *J. Climate*, **31**, 7481–7493, https://doi.org/10.1175/JCLI-D-18-0081.1.

St. George, S., and T. R. Ault, 2014: The imprint of climate within Northern Hemisphere trees. *Quat. Sci. Rev.*, **89**, 1–4, https://doi.org/10.1016/j.quascirev.2014.01.007.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Tolwinski-Ward, S. E., M. N. Evans, M. K. Hughes, and K. J. Anchukaitis, 2011: An efficient forward model of the climate controls on interannual variation in tree-ring width. *Climate Dyn.*, **36**, 2419–2439, https://doi.org/10.1007/s00382-010-0945-5.

Welch, P. D., 1967: The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.*, **15**, 70–73, https://doi.org/10.1109/TAU.1967.1161901.

Zhu, F., and Coauthors, 2019: Climate models can correctly simulate the continuum of global-average temperature variability. *Proc. Natl. Acad. Sci. USA*, **116**, 8728–8733, https://doi.org/10.1073/pnas.1809959116.