Separating internal and externally forced contributions to global temperature variability using a Bayesian stochastic energy balance framework

Cite as: Chaos **32**, 113146 (2022); https://doi.org/10.1063/5.0106123 Submitted: 27 June 2022 • Accepted: 01 November 2022 • Published Online: 29 November 2022

🔟 Maybritt Schillinger, 🔟 Beatrice Ellerhoff, ២ Robert Scheichl, et al.

COLLECTIONS

Paper published as part of the special topic on Theory-informed and Data-driven Approaches to Advance Climate Sciences





ARTICLES YOU MAY BE INTERESTED IN

Guidelines for data-driven approaches to study transitions in multiscale systems: The case of Lyapunov vectors

Chaos: An Interdisciplinary Journal of Nonlinear Science **32**, 113145 (2022); https://doi.org/10.1063/5.0093804

Finite-time recurrence analysis of chaotic trajectories in Hamiltonian systems Chaos: An Interdisciplinary Journal of Nonlinear Science **32**, 113144 (2022); https:// doi.org/10.1063/5.0102424

Chaotic synchronization of mutually coupled systems-arbitrary proportional linear relations Chaos: An Interdisciplinary Journal of Nonlinear Science **32**, 113137 (2022); https:// doi.org/10.1063/5.0100897



Chaos

Special Topic: Nonlinear Model Reduction From Equations and Data

Submit Today!



Chaos **32**, 113146 (2022); https://doi.org/10.1063/5.0106123 © 2022 Author(s).

ARTICLE

w Onlin

Separating internal and externally forced contributions to global temperature variability using a Bayesian stochastic energy balance framework

Cite as: Chaos **32**, 113146 (2022); doi: 10.1063/5.0106123 Submitted: 27 June 2022 · Accepted: 1 November 2022 · Published Online: 29 November 2022

Maybritt Schillinger,¹ b Beatrice Ellerhoff,^{2,a)} Beatrice Scheichl,³ hand Kira Rehfeld²

AFFILIATIONS

¹Seminar for Statistics, Department of Mathematics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland
 ²Department of Physics and Department of Geosciences, Tübingen University, Schnarrenbergstr. 94-96, 72076 Tübingen, Germany

³Institute of Applied Mathematics and Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

Note: This article is part of the Focus Issue, Theory-informed and Data-driven Approaches to Advance Climate Sciences. ^{a)}Author to whom correspondence should be addressed: beatrice-marie.ellerhoff@uni-tuebingen.de

ABSTRACT

Earth's temperature variability can be partitioned into internal and externally forced components. Yet, underlying mechanisms and their relative contributions remain insufficiently understood, especially on decadal to centennial timescales. Important reasons for this are difficulties in isolating internal and externally forced variability. Here, we provide a physically motivated emulation of global mean surface temperature (GMST) variability, which allows for the separation of internal and external variations. To this end, we introduce the "ClimBayes" software package, which infers climate parameters from a stochastic energy balance model (EBM) with a Bayesian approach. We apply our method to GMST data from temperature observations and 20 last millennium simulations from climate models of intermediate to high complexity. This yields the best estimates of the EBM's *forced* and *forced* + *internal* response, which we refer to as emulated variability. The timescale-dependent variance is obtained from spectral analysis. In particular, we contrast the emulated *forced* and *forced* + *internal* variance on interannual to centennial timescales with that of the GMST *target*. Our findings show that a stochastic EBM closely approximates the power spectrum and timescale-dependent variance of GMST as simulated by modern climate models. Small deviations at interannual timescales can be attributed to the simplified representation of internal variability and, in particular, the absence of (pseudo-)oscillatory modes in the stochastic EBM. Altogether, we demonstrate the potential of combining Bayesian inference with conceptual climate models to emulate statistics of climate variables across timescales.

© 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0106123

Understanding the statistical properties and sources of the Earth's surface temperature variations is of great importance in climate science. To this end, we analyze the variability of global mean surface temperature (GMST) with a simple stochastic energy balance model (EBM). With Bayesian methods and spectral analysis, we separate internally generated and externally forced contributions to GMST variations on different timescales in state-of-the-art climate model simulations. Our results show that a stochastic EBM can emulate the variability of more

complex climate models. The combined use of Bayesian inference and conceptual climate models, therefore, provides a versatile tool to advance the understanding of the internal and forced variability in the Earth's dynamical system.

I. INTRODUCTION

Climate variability describes the spatial and temporal variations in the mean and higher order statistics of climate parameters and is of vital importance for living conditions on the Earth.¹ While many sources of natural variability exist, anthropogenic influences clearly dominate the recent trend in global mean surface temperature (GMST). To characterize variability, it is typically partitioned into internal and external components. Internal variability arises from intrinsic climate system processes such as oceanic and atmospheric circulation. External sources include changes in radiative forcing, for example, from solar irradiance, volcanic eruptions, and greenhouse gases. Despite a general agreement of the total simulated and observed GMST variability over the Common Era (0–2000 CE),^{2,3} uncertainties remain about the mechanisms and magnitude of internal and external variations,^{4–6} especially on decadal to centennial timescales.^{3,7}

Simple mathematical models help understand climate variability⁸⁻¹⁰ and can be used to emulate climate variables from more complex model simulations. Most general, the time evolution of a forced climate parameter X(t) is described by $\dot{X}(t) = \mathscr{A}(t, X(t), F(t))$ for an arbitrary operator \mathscr{A} and external driver F(t). We consider X as the GMST, for which many studies have formulated physically motivated approximations of \mathscr{A} . One pivotal approach is centered around the idea of balancing incoming and outgoing radiations,^{8,9} later extended to a stochastic energy balance model (EBM) by Hasselmann.¹⁰ This approach assumes the climate system close to equilibrium, showing a linear and stationary response to perturbations. Then, \mathscr{A} can be approximated by a linear stochastic operator^{8,9}

$$C\frac{\mathrm{d}}{\mathrm{d}t}T(t) = -\tilde{\lambda}T(t) + F(t) + \varepsilon(t). \tag{1}$$

Formula (1) describes the GMST anomaly T(t) with respect to the equilibrium state, given the Earth's effective heat capacity C, a radiative forcing anomaly F(t), and a term $\varepsilon(t)$, representing stochastic dynamics such as weather fluctuations. The response parameter $\lambda := \tilde{\lambda}/C$ is the reciprocal of the characteristic timescale $1/\lambda$. The response to radiative forcing F(t) determines forced temperature variations. The response to the stochastic term $\varepsilon(t)$ approximates internal variability.

The stochastic EBM (1) is too simplistic to accurately represent long-term responses and, therefore, has been extended to the so-called multibox EBMs.^{11–15} The latter are based on multiple ocean layers, referred to as boxes. The layers serve to approximate the vertical heat transfer and the integrated response to forcing over long periods. The EBM (1) laid the basis for attributing anthropogenic warming.^{16,17} It was applied and modified to study climate sensitivity,^{14,18-20} climate and ice cap stability,²¹⁻²⁴ regional temperatures,^{25–27} glacial/interglacial cycles,^{22,28} and future projections.^{29,30} Key advantages of EBMs are their computational efficiency and comparatively easy interpretation.

To estimate uncertain parameters of conceptual climate models from data, Bayesian frameworks have become increasingly popular.³¹⁻³⁵ In comparison to other methods for inferring climate parameters, such as maximum likelihood estimation, Bayesian approaches have the advantage of providing full posterior distributions. The methods compute the posterior means and credible intervals (CIs) of uncertain parameters θ conditioned on target data *y* while including prior knowledge on θ . Central to this framework is applying the Bayes theorem

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)},$$
(2)

with likelihood $p(y|\theta)$, prior $p(\theta)$, marginal p(y), and posterior $p(\theta|y)$. We denote all probability densities by p and distinguish them by their arguments. Combining Bayesian inference with conceptual climate models typically also yields the posterior of the model's fit to the data.

With the ability to quantify fluctuations across timescales, power spectral analysis has improved the understanding of climate variability.^{3,36–39} Simple climate models have been combined with spectral analysis to explain timescale-dependent variability. For example, Fredriksen and Rypdal¹⁵ use a multibox EBM to study temporal scaling of temperature time series. Related works examine future projections²⁹ and climate sensitivity.⁴⁰ Soldatenko and Colman⁴¹ study the sensitivity of the power spectrum on uncertainties in the parameters of a two-box EBM, considering stochastic noise but neglecting deterministic forcing. Yet, the potential of combining stochastic multibox EBMs, Bayesian inference and spectral analysis to study the magnitude of unforced and forced variability across timescales remains untapped.

Here, we examine and separate timescale-dependent internal and externally forced contributions to the GMST variations. In particular, we analyze GMST variability during the last millennium (850-1850 CE) as simulated by 20 climate models of intermediate to high complexity. To this end, we combine a stochastic two-box EBM (Sec. III A) with Bayesian inference (Sec. III B) and spectral analysis (Sec. III C). We present the "ClimBayes" software package42 for Bayesian inference of climate parameters, which fits the stochastic EBM to GMST data. This results in the best estimate of the forced and samples of the *forced* + *internal* EBM's temperature response. First, we demonstrate our analysis on the example of historical observations (Sec. IV A) and then apply it to the considered set of last millennium simulations (Sec. IV B). Section IV C contrasts power spectra of the fitted EBM with and without internal variations. Comparing the internal and forced variance on interannual to centennial timescales (Sec. IV D), a stochastic two-box EBM captures most variations of more comprehensive model simulations. We summarize and discuss the potential for physics-informed emulation of GMST data and separation of variance contributions across timescales in Secs. V and VI.

II. DATA

Our study relies on annual GMST and corresponding radiative forcing time series. We use full-forced last millennium runs from climate models of varying complexity (Table I). We analyze 10 simulations with atmosphere-ocean general circulation models (AOGCMs), considered in the Coupled Model Intercomparison Project 5 (CMIP5).⁴³ Moreover, we use 10 simulations with Earth system models of intermediate complexity (EMICs) that are part of the IPCC's Fifth Assessment Report (AR5)⁴⁴ and described by Eby *et al.*⁴⁵ The AR5 EMICs represent single simulations, except for CLIMBER2 and LOVECLIM V.1.2, which are ensemble means and

ARTICLE

TABLE I. Key forcing specifications and references of considered climate model simulations. The "Forcing" column gives the abbreviations from the PMIP3 protocol,⁴⁶ corresponding to the implemented land use, solar, and volcanic forcing reconstructions. The land use reconstruction PEA is taken from Pongratz *et al.*⁴⁹ Solar forcing reconstructions correspond to DB: Delaygue and Bard,⁵³ VSK: Krivova, Balmaceda, and Solanki,⁵⁴ Vieira and Solanki,⁵⁵ and SBF: Steinhilber, Beer, and Fröhlich.⁵¹ They are calibrated to WLS modern values (1366.14 W/m²) and continued by Wang, Lean, and Sheeley.⁵² Volcanic forcing refers to CEA: Crowley *et al.*⁵⁰ and GRA: Gao, Robock, and Ammann.⁵⁶ Trace gases are prescribed in all simulations and follow the PMIP3 protocol.⁴⁶

Climate model	nate model Forcing	
AR5 EMICs		
Bern 3D	PEA, DB, CEA	57
CLIMBER-3alpha	PEA, DB, CEA	58
CLIMBER2	PEA, DB, CEA	59
DCESS ESM v1	PEA, DB, CEA	60
IGSM 2.2	PEA, DB, CEA	61
LOVECLIM V.1.2	PEA, DB, CEA	62
MESMO 1.0	PEA, DB, CEA	63
MIROC3-lite	PEA, DB, CEA	47
UMD	PEA, DB, CEA	64
UVic v2.9	PEA, DB, CEA	65
CMIP5 models		
BBC-CSM1-1	–, VSK, GRA	66
CCSM4	PEA, VSK, GRA	67
CSIRO-Mk3L-1-2	–, SBF, CEA	68
FGOALS-s2	–, VSK, GRA	69
GISS-E2-R	PEA, SBF, CEA	70
HadCM3	PEA, SBF, CEA	71
HadGEM2-ES	PEA, SBF, CEA	72
IPSL-CM5A-LR	–, VSK, GRA	73 and 74
MIROC-ESM	–, DB, CEA	75
MPI-ESM-P	PEA, VSK, CEA	76 and 77

denoted by "(mean)" in the following. To compare variability in single ensemble members to that of the ensemble mean, we use the five available ensemble members LOVECLIM V.1.2 (E1–E5).

The transient radiative forcing applied to these simulations follows the Paleoclimate Modelling Intercomparison Project Phase III (PMIP3) protocol.⁴⁶ For AR5 EMICs, we take the total estimated radiative forcing provided by Eby et al.⁴⁵ For CMIP5 simulations, we use the radiative forcing from reconstructions of well-mixed greenhouse gases (CO₂, CH₄, and N₂O), volcanic aerosols, total solar irradiance, and land use changes as provided by Schmidt et al.⁴⁶ We neglect orbital forcing, which is assumed to play a negligible role for GMST variability over the last millennium. To remove potential unforced drifts of the simulated background climate,⁴⁷ the simulated GMST is linearly detrended prior to analysis. For consistency, detrending is also applied to the corresponding forcing time series. This does not affect our results, as the simulations' forcing input exhibits no trend for the last millennium (850-1850 CE). Both temperature and forcing time series are considered as anomalies with respect to the starting year.

We use the GMST from HadCRUT5⁴⁸ observations (1850–2000 CE) to demonstrate the developed workflow of our Bayesian stochastic energy balance framework. As estimates for radiative forcing during the historical period, we consider the "PEA" land use (Pongratz *et al.*⁴⁹), "CEA" volcanic forcing (Crowley *et al.*⁵⁰), "SBF" solar irradiance reconstruction (Steinhilber *et al.*⁵¹) patched into Wang *et al.*⁵² and greenhouse gas concentrations from Schmidt *et al.*⁴⁶

III. METHODS

Our analysis combines stochastic multibox EBMs, Bayesian inference, and spectral analysis. We introduce the approach implemented in "ClimBayes"⁴² for the most generic case of a stochastic EBM with *N* boxes in Sec. III A. All results are obtained from the special case N = 2.

A. Stochastic two-box energy balance model (EBM)

The stochastic multibox EBM^{13,15,29} extends the onedimensional linear operator from Eq. (1) by multiple vertical layers, approximating the heat exchange between surface and deep ocean layers. In matrix notation, the model reads¹⁵

$$C\frac{dT}{dt}(t) = KT(t) + F(t) + \varepsilon(t).$$
(3)

For *N* boxes, T(t) is an *N*-dimensional vector, describing the temperature of each box. By convention, T_1 corresponds to the temperature of the uppermost and T_N to the temperature of the lowermost box. Accordingly, *C* is a diagonal matrix with the effective heat capacity C_{ii} of each layer (i = 1, ..., N). *K* is a *N*-dimensional tridiagonal matrix, parameterizing the surface temperature response and vertical heat transfer (Appendix A). The time-dependent radiative forcing F(t) is only applied to the uppermost box, such that $F_1 = F(t)$ and $F_k = 0$ for k = 2, ..., N. The stochastic forcing $\varepsilon(t)$ is likewise implemented with non-zero entry $\varepsilon_1(t) = \sigma_W \xi(t)$. We motivate the white noise process $\xi(t)$ with standard deviation (SD) σ_W by the found impact of weather fluctuations.^{10,14,78}

Integrating Eq. (3) yields the solution of the surface temperature $T_1(t)$, given by a sum of the forced response $T_{1,F}(t)$ and internal variations $T_{1,I}(t)$,

$$T_{1}(t) = T_{1,F}(t) + T_{1,I}(t)$$

= $\int_{-\infty}^{t} R(t-s) \frac{1}{C_{1}} F(s) \, ds + \int_{-\infty}^{t} R(t-s) \frac{\sigma_{W}}{C_{1}} \, dW(s),$ (4)

where $C_1 = C_{11}$ is the heat capacity of the uppermost box. This assumes no interaction between forced and internal variability on a global scale. The response function

$$R(t) = \sum_{k=1}^{N} w_k e^{-\lambda_k t}$$
(5)

is uniquely defined^{13,15} by the response parameters λ_k and weights $w_k \ (k = 1, ..., N)$ with $\sum_{k=1}^{N} w_k = 1$, which depend on the entries of C and K (Appendix A). The internal fluctuations $T_{1,I}(t)$ in formula (4) represent an Itô-integral over the Wiener process W(s). Therefore, $T_{1,I}(t)$ can be written as a weighted sum of Ornstein–Uhlenbeck

(OU) processes, where the *k*th OU process solves the stochastic differential equation $dU(t) = -\lambda_k U(t) dt + \frac{\sigma_W}{C_1} dW(t)$ and receives weight w_k . Accordingly, $T_{1,I}(t)$ is normally distributed with mean zero. Its covariance matrix is determined by σ_W , C_1 , w_k , and λ_k (Appendix B).

B. Bayesian inference algorithm

1. Joint emulation of forced and internal variations

To separate the internal and forced contributions to the GMST variance, we introduce the Bayesian inference algorithm implemented in "ClimBayes." We fit the linear stochastic two-box EBM to GMST data (Table I), as illustrated in Fig. 1. The Bayesian inference algorithm relies on annually resolved temperature and forcing time series as input data. Moreover, it requires physics-informed prior information on $\theta = (\lambda_1, \lambda_2, w_1, C_1, T_0, F_0)$. The parameters λ_1, λ_2 , and w_1 correspond to free parameters of the response function in Eq. (5). C_1 is the heat capacity of the upper ocean box. T_0 and F_0 are initial free parameters. T_0 allows for small deviations of the EBM solution to the input data in the starting year and is expected to be close to zero. Similarly, the additional parameter F_0 is needed to compensate for an initial forcing anomaly with respect to the equilibrium state.

We infer the posterior distributions of the uncertain parameters θ conditioned on target data *y* via Bayes theorem (2), using a Markov chain Monte Carlo (MCMC) algorithm. To this end, we assume that the target data can be described by a deterministic model Φ_F and stochastic measurement or intrinsic noise *Z* that is also allowed to depend on θ ,

$$y = \Phi_F(\theta) + Z(\theta). \tag{6}$$

Formula (6) yields the likelihood $p(y|\theta)$. Combined with prior information $p(\theta)$, Bayes theorem (2) defines the posterior distribution $p(\theta|y)$. In our case, the deterministic model $\Phi_F(\theta)$ is given by a discretization of the temperature responses $T_{1,F}(t)$. The noise term $Z(\theta)$ corresponds to the internal fluctuations $T_{1,I}(t)$.

This approach provides a joint estimate of the internal and externally forced response, based on the same physics-informed response function and parameters. The best estimates of θ and the forced response $T_{1,F}(t)$ are defined as their posterior means $E[\theta|y]$ and $E[\Phi_F(\theta)|y]$. The SD σ_I of the internal variability $Z(\theta)$ determines σ_W (Appendix B) and is approximated by the residuals' SD, that is, the difference between the target data and the forced response. Samples of the internal variations $T_{1,I}(t)$ can be drawn from its covariance matrix, using the best estimates of θ (Appendix B). The forced + internal variations $T_{1,F}(t) + T_{1,I}(t)$ represent the full response of the stochastic two-box EBM and, thus, provide a model for the target variability. To streamline our discussion, we will refer to the modeled forced and forced + internal response as emulation.

2. Numerical implementation

The "ClimBayes" package provides the numerical implementation of our approach (Fig. 1) and allows for straightforward adjustments via a configuration file. Most importantly, this includes specifications of the number of boxes, prior distributions, MCMC

$$\mathbf{C}\frac{d\mathbf{I}}{dt}(t) = \mathbf{K}\mathbf{T}(t) + \mathbf{F}(t) + \boldsymbol{\epsilon}(t)$$

$$T_{1}(t) = \int_{-\infty}^{t} R(t-s) \frac{1}{C_{1}} F(s) ds + \int_{-\infty}^{t} \frac{R(t-s)}{C_{1}} \frac{\sigma_{W}}{dW(s)} dW(s)$$

$$R(t) = w_{1}e^{-\lambda_{1}t} + w_{2}e^{-\lambda_{2}t}$$

b Data

Temperature observations y and forcing F(t)Prior $p(\theta)$ of uncertain variables $\theta = (\lambda_1, \lambda_2, w_1, T_0, F_0)$

- Bayesian inference with MCMC algorithm

Assumption $y = \Phi_F(\theta) + Z(\theta) = T_1(t)$ *observations forced internal* Bayes theorem $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$

^d → Results

Posterior
$$p(\theta|y)$$

Best Bayesian estimate of forced and internal variations

FIG. 1. Workflow of our Bayesian inference algorithm to emulate forced and internal GMST fluctuations. (a) The workflow builds on a linear stochastic two-box EBM, here in matrix notation. The solution for the surface temperature $T_1(t)$ is given by an integral with exponential response function R(t). (b) Required input data include annually resolved GMST and time-dependent forcing, as well as physics-informed prior information $p(\theta)$ on uncertain parameters θ . (c) We infer the parameters of the two-box EBM using a Markov chain Monte Carlo (MCMC) algorithm and Bayes theorem, assuming that the target GMST can be described as a deterministic model $\Phi_F(\theta)$ and stochastic noise $Z(\theta)$. (d) The workflow yields posterior distributions $p(\theta|y)$ of uncertain parameters conditioned on temperature data *y* and a physically motivated emulation of the forced and internal variations.

sampling properties, and fixed parameters. We choose N = 2 boxes (Appendix C) in line with Held *et al.*¹² and Geoffroy *et al.*¹³ Our experiments use independent prior distributions. We choose beta distributions with shape parameters $\alpha = \beta = 2$ for the marginal priors of λ_1 , λ_2 , as well as the initial parameters T_0 and F_0 . This allows for a physics-informed, fixed parameter range, where the mean is preferred over boundary values. The algorithm's convergence is improved compared to uniform priors. The intervals for the response parameters $\lambda_1 : (0.2, 2) \text{ yr}^{-1}$ and $\lambda_2 : (0.005, 0.2) \text{ yr}^{-1}$ are similar to those from Fredriksen and Rypdal.¹⁵ Tailored to our goal to emulate interannual to centennial variability from last millennium simulations, our choice of priors assumes characteristic response times $1/\lambda_k$ smaller than 200 years. These response times implicitly set a characteristic depth scale for the two ocean boxes of our stochastic EBM.

The prior of C_1 : (4, 11) W yr m⁻² K⁻¹ follows previous findings.^{13,15,79} Initial values T_0 : (-0.5, 0.5) K and F_0 : (-2, 2) W m⁻² are centered around zero. The weight w_1 : (0, 1) has an uniform prior. We consider no measurement noise, which is inherently

fulfilled for simulated GMST. For observed GMST, we assume measurement errors to be small compared to internal fluctuations.⁴⁸ We verified that our findings are robust against reasonable variations of the MCMC and prior specifications.

To obtain the best parameter estimates, "ClimBayes" uses a Metropolis Hastings (MH) algorithm from the family of MCMC methods. To this end, we discretize the forward operator and compute annual temperature anomalies relative to the starting point t = 0 by a midpoint rule²⁹

$$T_{1,F}(t_m) = T_0 + \frac{1}{C_1} \sum_{k=1}^{N} w_k \sum_{j=1}^{m} e^{-(m-j+1/2)\lambda_k} (F(t_j) + F_0).$$
(7)

Here, the time step t_j corresponds to the *j*th year, and $F(t_j)$ are forcing anomalies with respect to the starting point t = 0. F_0 represents an initial forcing anomaly. Discretizing $T_{1,I}$ leads to a normally distributed weighted sum of AR(1)-processes with covariance matrix entries depending on λ_k and w_k (Appendix B). The likelihood $p(y|\theta)$ is given by the normal distribution of $T_{1,F}(t) + T_{1,I}(t)$. This requires the calculation of the covariance matrix for each sample in the Markov chain. For numerical robustness and computational efficiency, however, we approximate the likelihood function using an iterative scheme (Appendix B).

The MCMC algorithm uses four chains with 25 000 samples each, from which the first 5000 are discarded as burn-in. The proposal distribution is initially set to a normal distribution with mean zero and variances $(0.2, 2, 1, 10, 1, 2) \times 10^{-5}$ for $\theta = (\lambda_1, \lambda_2, w_1, C_1, T_0, F_0)$. After 2500 samples, the proposals are distributed according to the weighted sum of the initial normal proposal distribution and the empirical covariance matrix of previous samples.

We check the convergence of the Markov chains following two performance measures: First, the Gelman–Rubin diagnostics^{80,81} compares the inter-chain and between-chain variances. It is ≤ 1.1 for most models, complying to recommendations.^{82,83} Second, we use the Monte Carlo standard error,⁸⁴ which constructs an asymptotic confidence interval for the posterior mean.^{84,85} In our experiments, the half-width of this interval is smaller than 5% of the prior mean. We found that this criterion guarantees robustness of the parameter estimates when the same run is repeated multiple times or additional samples are added. DCESS ESM v1 is the only outlier, showing a tendency for bimodal posterior densities which lead to less stable estimates with wide CIs. However, we confirmed that convergence can be achieved by fixing the heat capacity to the estimated parameter.

C. Spectral analysis and variance ratios

Given a temperature time series T(t), the power spectral density (PSD) at frequency f corresponds to the Fourier transform of the autocovariance

$$S(f) = \int_{-\infty}^{\infty} e^{-2\pi i f k} E\left[(T(t) - \mu) \left(T(t+k) - \mu \right) \right] dk, \quad (8)$$

with lag $k = t_2 - t_1$ and mean $\mu := E[T(t)]$. This assumes X(t) to be weakly stationary, which is reasonably fulfilled after linear detrending the GMST data. Following Ellerhoff and Rehfeld,³ we use the

multitaper method with three windows to compute the PSD and χ^2 -distributed uncertainties. Mean spectra are obtained after interpolation to the lowest resolution and binning into equally spaced log-frequency intervals.⁸⁶ The spectra are visualized over periods $\tau = 1/f$ and logarithmically smoothed using a Gaussian kernel of 0.04 dB.

We compare the PSD for three types of time series: (1) the *tar-get* temperature data from historical observations or climate model simulations, AR5 EMIC and CMIP5 models, (2) the emulated *forced* variations $T_{1,F}(t)$, and (3) the emulated *forced* + *internal* temperature variations $T_{1,F}(t) + T_{1,I}(t)$ from the stochastic two-box EBM. To compute the *forced* + *internal* PSD, we first sample 1000 realizations of the internal response $T_{1,I}(t)$. We add these to $T_{1,F}(t)$ and compute the PSD for all samples. Mean spectra and 95% confidence bands are obtained from this ensemble.

Variance ratios are calculated by dividing the emulated by the *target* variance. Following Parseval's theorem, we determine the timescale-dependent variance by integrating the PSD over frequencies. We consider frequency bands corresponding to interannual (2-5 yr), decadal (5-20 yr), multidecadal (20-50 yr), and centennial (50-200 yr) scales.

IV. RESULTS

A. Example application to historical observations

We demonstrate the application of the Bayesian inference algorithm and our spectral analysis on the example of GMST observations from HadCRUT5.⁴⁸ Figure 2(a) shows the forcing and temperature time series together with the best estimate of the forced response $T_{1,F}(t)$. The forced response follows the global warming trend and shows cooling periods after volcanic eruptions. Credible intervals (CIs) capture the uncertainties of the forced response, but not those due to internal variability. As a result, observations partly lie outside CIs. Uncertainties of the two-box forced response are largest at the time series' start.

Figure 2(b) shows marginal prior and posterior distributions for the free parameters of the stochastic two-box EBM. The response parameters $\lambda_1 = 1.31 (0.71, 1.86) \text{ yr}^{-1}$ and $\lambda_2 = 0.09 (0.03, 0.16) \text{ yr}^{-1}$ (Table III) correspond to timescales of approximately 10 months and 10 years. The corresponding heat capacity C_1 of the upper ocean layer is 8.20 (5.68, 10.43) W yr m⁻² K⁻¹. The initial values T_0 and F_0 are well constrained and close to zero. The weight $w_1 = 0.72 (0.46, 0.90)$ tends to emphasize the fast response.

The power spectral density [Fig. 2(c)] of the *forced* response alone is smaller than that of the *target* temperature. Conversely, the magnitude of the emulated *forced* + *internal* PSD agrees with the *target* PSD within uncertainties, except for the interannual scale. While the emulated PSD constantly increases from interannual to multidecadal scales, HadCRUT5 shows a modulation with increased power on periods of two to ten years, which is not captured by the emulated response.

Figure 2(d) compares the variance on interannual to multidecadal timescales. The variance ratios are formed by dividing the emulated *forced* or *forced* + *internal* variance by the variance obtained from the HadCRUT5 *target*. The *forced* variance is smaller than the *target* variance on all timescales. Incorporating internal



FIG. 2. Example application of our developed approach to historical GMST observations from HadCRUT5. (a) *Target* data *y*, forcing *F*(*t*) (taken from Schmidt *et al.*,⁴⁶ Pongratz *et al.*,⁴⁰ Crowley *et al.*,⁵⁰ Steinhilber *et al.*,⁵¹ Wang *et al.*⁵¹ and the best estimate of the *forced* response $T_{1,F}(t)$, that is the posterior mean $E[\Phi_F(\theta)|y]$. Shaded areas correspond to 95% Cls of $T_{1,F}(t)$. (b) Marginal prior and posterior densities for uncertain parameters θ . (c) PSD of the GMST observations, the *forced* response (both with χ^2 -distributed confidence bands) and the sampled *forced* + *internal* variations. The sampled *forced* + *internal* PSD represents the mean and 95% confidence bands obtained from an ensemble of $T_{1,F}(t) + T_{1,I}(t)$ using 1000 realizations of $T_{1,I}(t)$. (d) Ratios of the emulated to observed GMST variance, computed by integration of the PSD on the multidecadal (20–100 yr), decadal (5–20 yr), and interannual (2–5 yr) scales. Uncertainties (95% Cl) are calculated from a F-distribution based on the degrees of freedom of the variance estimate.

variability reduces this mismatch strongly, yet, is not enough to capture all fluctuations on interannual and multidecadal scales. On decadal scales, the emulated *forced* + *internal* variability agrees well with the observations.

B. Parameters estimated from last millennium simulations

We use Bayesian inference to fit the stochastic two-box EBM to GMST simulations from CMIP5 models and AR5 EMICs (Sec. II). Table II shows the best estimates that are the posterior means and 95% CIs of $\theta = (\lambda_1, \lambda_2, w_1, C_1, T_0, F_0)$ as well as the SD σ_I of the internal variability $T_{1,I}(t)$. Across all simulations, the short-term response parameter λ_1 varies between 1.91 and 0.31 yr⁻¹, spanning the full prior range between 6 months and 5 years. The long-term response λ_2 varies between 0.19 and 0.01 yr⁻¹, corresponding to characteristic timescales of approximately 5–100 years.

Differences between AR5 EMICs and CMIP5 models are most pronounced for λ_1 . CMIP5 simulations exhibit larger CIs and intermodel differences, while AR5 EMICs show similar λ_1 , except for CLIMBER 2 (mean) and LOVECLIM (E1 and mean). The weight $w_1 > 0.5$ is typically larger than $w_2 = 1 - w_1$, emphasizing the relative importance of the fast compared to the slow response. w_1 often is closer to unity for AR5 EMICs than for CMIP5 models. CLIMBER2 (mean) shows exceptionally large λ_1 and small w_1 . The heat capacity C_1 varies from 4.7 to 10.96 W yr m⁻¹ K⁻¹. It spans the full prior range for both AR5 EMICs as well as CMIP5 models and shows no strong dependence on other parameters. The initial temperature T_0 is well constrained and close to zero. Inter-model differences are also large for F_0 and linked to varying temperature amplitudes at the beginning of the time series with respect to the mean (Fig. 7). The

TABLE II. Posterior means and 95% CIs of uncertain parameters θ for the stochastic two-box EBM fitted to GMST from climate model simulations. The SD σ_I of the internal
variability equals the residual of the forced response (Appendix B) and is, therefore, given without CIs. Ensemble means are denoted by "(mean)." We exemplary show the
LOVECLIM ensemble member "E1" as there are no major differences across the ensemble. Models with GRA56 volcanic forcing are marked with an asterisk (*), all others use
the CEA reconstruction ⁵⁰ (Table I). The weight $w_2 = 1 - w_1$ (not shown) is uniquely defined by w_1 .

Climate	(-1)	(m^{-1})	w (writing)	C_{1} (MI	T(V)	$E_{\rm c}$ (M/m ⁻²)	-(V)
	λ_1 (yr)	$\lambda_2 (yr)$	w_1 (unitiess)	C_1 (w yr m K)	$I_0(\mathbf{K})$	F_0 (WIII)	$O_I(\mathbf{K})$
AR5 EMICs Bern 3D	0 74 (0 62 0 81)	0.08 (0.06.0.15)	0 91 (0 88 0 95)	4 76 (4 54 5 43)	-0.09(-0.500.04)	0.18(-0.15, 1.48)	0.02
CLIMBER-	0.71 (0.02,0.01)	0.00 (0.00,0.10)	0.91 (0.00,0.90)	1.70 (1.51, 5.15)	0.09 (0.00,0.01)	0.10 (0.10, 1.10)	0.02
3alpha	0.76 (0.72,0.80)	0.06 (0.05,0.07)	0.91 (0.90,0.93)	6.18 (5.99, 6.35)	-0.02 (-0.05,0.02)	-0.04 (-0.13, 0.04)	0.01
CLIMBER2							
(mean)	1.43 (0.25,1.97)	0.19 (0.17,0.20)	0.05 (0.00,0.19)	10.96 (10.89,11.00)	0.01 (-0.11,0.13)	-0.19(-0.45, 0.07)	0.05
DCESS ESM							
v1	0.31 (0.21,1.18)	0.08 (0.03,0.18)	0.64 (0.00,0.92)	10.25 (9.95,10.52)	0.02 (-0.05,0.08)	-0.18 (-0.27,-0.05)	0.03
IGSM 2.2	0.64 (0.57,0.73)	0.10 (0.07,0.15)	0.89 (0.82,0.93)	6.68 (6.38, 6.98)	0.01 (-0.08,0.11)	-0.10 (-0.36, 0.15)	0.03
LOVECLIM							
(E1)	1.32 (0.99,1.69)	0.13 (0.07,0.18)	0.90 (0.85,0.94)	6.03 (4.95,7.29)	-0.01 (-0.22,0.19)	0.28 (-0.58,1.21)	0.09
LOVECLIM							
(mean)	1.12 (0.93,1.34)	0.14 (0.09,0.19)	0.88 (0.84,0.92)	6.74 (6.06,7.42)	0.04 (-0.08,0.16)	-0.50(-1.03,0.01)	0.04
MESMO 1.0	0.52 (0.49,0.56)	0.03 (0.02,0.04)	0.90 (0.88,0.92)	7.96 (7.72, 8.19)	0.02 (-0.03,0.06)	-0.32 (-0.39,-0.24)	0.02
MIROC-lite	0.44 (0.41,0.47)	0.02 (0.01,0.04)	0.96 (0.94,0.97)	6.75 (6.56, 6.96)	0.01 (-0.05,0.08)	-0.34 (-0.49,-0.25)	0.03
UMD	0.72 (0.68,0.77)	0.01 (0.01,0.01)	0.99 (0.99,0.99)	8.81 (8.46, 9.15)	0.00 (-0.04,0.02)	-0.09 (-0.21, 0.08)	0.02
UVic v2.9	0.45 (0.41,0.51)	0.05 (0.03,0.08)	0.88 (0.81,0.93)	10.88 (10.68,10.98)	-0.01 (-0.06,0.05)	-0.05 (-0.19, 0.09)	0.02
CMIP5 models BCC-							
CSM1-1*	1.91 (1.76,1.99)	0.06 (0.04,0.10)	0.94 (0.91,0.96)	10.49 (9.78,10.92)	-0.05 (-0.13,0.06)	1.31 (0.53, 1.87)	0.10
CCSM4*	1.62 (1.27,1.90)	0.12 (0.08,0.16)	0.88 (0.83,0.92)	4.70 (4.12, 5.60)	-0.06 (-0.34,0.18)	0.17 (-0.60, 1.05)	0.14
CSIRO-							
Mk2L-1-2	0.48 (0.29,1.12)	0.15 (0.07,0.19)	0.65 (0.31,0.96)	10.49 (9.78,10.92)	0.02 (-0.17,0.21)	-0.23 (-0.76, 0.30)	0.08
FGOALS-s2*	1.49 (1.13,1.88)	0.03 (0.01,0.10)	0.94 (0.89,0.97)	8.41 (6.71,10.38)	-0.05 (-0.20,0.11)	-0.58 (-1.42,-0.06)	0.14
GISS-E2-R	0.56 (0.50,0.65)	0.03 (0.01,0.15)	0.99 (0.96,1.00)	6.53 (5.94, 7.07)	0.12 (-0.43,0.32)	-1.19 (-1.77, 0.48)	0.09
HadCM3	0.49 (0.34,0.73)	0.13 (0.05,0.19)	0.76 (0.52,0.96)	7.99 (7.15, 8.86)	-0.06 (-0.31,0.20)	0.89 (0.27, 1.48)	0.12
HadGEM2-ES	0.64 (0.39,1.28)	0.14 (0.05,0.19)	0.74 (0.53,0.95)	7.79 (6.52, 8.81)	-0.06 (-0.30,0.19)	0.97 (0.36, 1.58)	0.12
IPSL-							
CM5A-LR*	1.69 (1.36,1.95)	0.09 (0.06,0.14)	0.85 (0.78,0.91)	9.49 (7.89,10.73)	-0.16 (-0.34,0.00)	-0.99 (-1.77,-0.14)	0.15
MIROC-							
ESM	1.49 (1.04,1.90)	0.14 (0.09,0.19)	0.71 (0.60,0.81)	9.62 (8.01,10.78)	-0.02 (-0.25,0.22)	0.10 (-0.83, 1.03)	0.11
MPI-ESM-P	0.66 (0.45,0.97)	0.14 (0.07,0.19)	0.77 (0.61,0.93)	6.11 (5.42, 6.80)	0.04 (-0.22,0.29)	-0.57 (-1.15,-0.01)	0.11

SD σ_I of the internal variability from CMIP5 models lies between 0.09 and 0.15 K and is larger than for AR5 EMICs (0.01–0.09 K).

C. Emulation of power spectral density

Figure 3 compares the *target* PSD to the emulated *forced* and *forced* + *internal* PSD. For most simulations, the emulated *forced* + *internal* PSD agrees with the *target* within uncertainties. AR5

EMICs show no major differences between *forced* and *forced* + *internal* PSD above decadal scales, except for the LOVECLIM ensemble members. Hence, the forced response is sufficient to emulate the long-term variability of most AR5 EMIC simulations. On interannual scales, considering internal variability in the emulation compensates for the mismatch between the *forced* and *target* PSD. Differences between emulated and *target* PSD are most pronounced for CLIMBER2 (mean), showing an overestimation on multidecadal to centennial scales by the emulation.



FIG. 3. Target and emulated forced and forced + internal power spectral density (PSD) and 95% confidence bands for AR5 EMICs [top: (a)–(k)] and CMIP5 models [bottom: (I)–(u)], as Fig. 1(c). Shaded intervals highlight examples of overestimation (pink) and underestimation (green) of the emulated PSD compared to the *target*, which are discussed in the main text. We show the PSD for LOVECLIM V.1.2 (mean) and the first ensemble member LOVECLIM V.1.2 (E1), as there are no major differences across ensemble members.

For CMIP5 simulations, the emulated *forced* PSD underestimates the *target* PSD on all timescales. Conversely, the *forced* + *internal* variability matches the *target* well for almost all models. Minor differences are found on interannual scales [Figs. 3(1) and 3(q)-3(u)]. Here, the *target* PSD of MPI-ESM-P and HadCM3 exhibit increased power on periods of 2–8 years. CCSM4, FGOALS-s2, and IPSL-CM5A-LR overestimate the PSD on the shortest timescales of approximately 2 years. The emulated *forced* + *internal* PSD for BCC-CSM1-1 deviates from that of the *target* by showing increased power on interannual and decreased power on multidecadal to centennial scales.

D. Separating internal and externally forced variance

Figure 4 shows the mean and spread of variance ratios on interannual to centennial scales for the considered model types. Variance ratios smaller than one indicate less emulated *forced* or *forced+internal* than *target* variance. We add a comparison of

the variance from the five LOVECLIM V.1.2 ensemble members (E1–E5) [Fig. 4(b)] to that of the remaining AR5 EMICs [Fig. 4(a)] and CMIP5 models [Fig. 4(c)]. Here, AR5 EMICs explicitly include the LOVECLIM ensemble mean, but not its members. For all model types, the relative contribution of internal variability decreases with increasing timescale, as the ratios for *forced* and *forced* + *internal* variance become more similar. The contribution of internal variability is larger in LOVECLIM ensemble members and CMIP5 simulations compared to AR5 EMICs.

The emulated variance of AR5 EMICs [Fig. 4(a)] is dominated by *forced* variations and matches the *target* variance on interannual and decadal scales. On longer timescales, the emulated variance tends to overestimate the *target* variance. This is mostly due to outliers, which correspond to CLIMBER2 (mean), in line with the overestimated PSD in Fig. 3(b). Members of the LOVECLIM V.1.2. ensemble [Fig. 4(b)] exhibit more internal contributions to the variance on all timescales compared to AR5 EMICs. The ensemble's emulated *forced* + *internal* variance approximates the *target* well on



FIG. 4. Ratios of the emulated forced and forced + internal to the target variance for AR5 EMICs [panel (a)], LOVECLIM ensemble members [panel (b)], and CMIP5 simulations [panel (c)] on centennial (50–200 yr), multidecadal (20–50 yr), decadal (5–20 yr), and interannual (2–5 yr) scales [as Fig. 1(d)]. Bars indicate the mean variance ratio over the considered simulations for one model type. Circles correspond to individual simulations. Confidence bands for individual ratios (as in Fig. 2) are not shown for better visibility.

the interannual and decadal scale. On larger timescales, we find an overestimation of the *target* variance.

The mean emulated *forced* + *internal* variance of CMIP5 simulations [Fig. 4(c)] is close to one on all timescales. The relative contribution of internal compared to forced variations is similar to that of LOVECLIM ensemble members on interannual and decadal timescales. However, there is a larger spread of variance ratios. Moreover, we find a small tendency of the emulation to overestimate the interannual and centennial variance of CMIP5 models. BCC-CSM1-1 represents an outlier, with the uppermost variance ratio on the shorter and lowermost on the longer timescales, in line with the spectral analysis [Fig. 3(1)].

V. DISCUSSION

We demonstrate the emulation of GMST variability as simulated by state-of-the-art climate models using a linear stochastic two-box EBM and Bayesian inference. Our analysis builds on the same, physically motivated response function for internal and external processes and allows for consistent separation of internal and externally forced variability. Estimates of the timescale-dependent variance show that the relative contribution of internal variability increases with model complexity and decreases with timescale.

Building on previous studies,^{29,31} the strength of our Bayesian framework is that it yields the posterior means and CIs for the uncertain parameters of the stochastic two-box EBM fitted to GMST simulations. Due to our choice of priors and the convergence of the fit, the estimated heat capacity C_1 agrees by construction with previous findings.^{13,15,79} Our response parameters λ_1 and λ_2 are consistent with results from Fredriksen and Rypdal,¹⁵ estimated from observational data. However, the estimates differ from those obtained in $4 \times CO_2$ experiments.^{13,79} This is because our framework accounts for high-frequency pulses and, thus, estimates response parameters associated with faster dynamics. Furthermore, our findings reveal a dependence of the estimated response parameters on the imprint of intermittent volcanic eruptions on simulated temperatures. This is reflected in consistently high values for w_1 , emphasizing the fast feedback. The inter-model spread of λ_1 in CMIP5 simulations suggests a link to the implemented volcanic forcing: CMIP5 simulations driven by a comparatively weak reconstruction ("GRA") tend to show higher values for λ_1 compared to those driven by "CEA" (Table I), which has greater forcing amplitudes. λ_1 and w_1 are particularly high for BCC-CSM1-1, indicating a fast and weak forced response of the fitted EBM (Fig. 7). This is consistent with a weak forced response in BCC-CSM1-1.87 The parameter estimates for CLIMBER2 (mean) differ from the remaining AR5 EMICs. In particular, λ_1 is poorly constrained, as CIs span the full prior range. We find that the temperature response to volcanic eruptions in CLIMBER2 (mean) is delayed. The estimated λ_1 and w_1 can be reconciled with those of the other AR5 EMICs if the temperature data were shifted by 1 yr. Altogether, the sensitivity of the fit to intermittent volcanic forcing suggests that high-frequency forcing plays a crucial role for simulating temperature variability across scales correctly.

Different methods have been developed to isolate forced and internal variations based on detrending,88 single-model ensembles,^{5,89-91} and deterministic EBMs,^{92,93} among others. The application of a Bayesian energy balance framework to the timescale-dependent quantification of forced and internal variance is novel. Our method provides a robust and joint separation of the variations at every step in time in a statistically sound way. Using data from the CESM Large Ensemble Community project,94 we verify the robustness of our *forced* response (Fig. 8). The latter agrees well with the ensemble mean, which shows higher variability due to remaining internal variations that have not been averaged out over the 13 members. Hence, our method provides a robust tool to estimate the *forced* response when large ensembles are not available. Moreover, we find a wide agreement of the emulated forced + internal variance with that from CMIP5 simulations. The fact that the stochastic two-box EBM mimics the temperature variations well is in line with previous findings on a linear relation between external forcing and GMST.^{3,13,15,95}

Small differences between the emulated forced+internal and target PSD for CMIP5 models on interannual timescales can be attributed to the simplified representation of internal variability as a weighted sum of AR(1) processes by the stochastic EBM. The latter cannot represent (pseudo-)oscillatory climate modes or modulations of internal variability by external forcing,96,97 which could result in the observed underestimation of the PSD by the emulation on these timescales. For HadCM3 and MPI-ESM [Figs. 3(s) and 3(u)], deviations on the interannual scale are similar to those in HadCRUT5 and likely due to the spectral imprint of the El Niño-Southern Oscillation.^{3,98} Moreover, our approach assumes that the covariance structure of the internal variations is determined by the estimated feedback parameters. In CMIP5 simulations with "GRA" volcanic forcing, particularly high values for the fast response parameter λ_1 and the weight w_1 lead to internal variations with autocorrelations on short timescales. This can cause an overestimation of the emulated PSD [Figs. 3(1), 3(q), 3(r), and 3(t)]. Similarly, the overestimation of the *target* PSD on longer timescales [Fig. 3(b)] in CLIMBER 2 can be explained by its estimated slow response. The latter is due to the biased delay between forcing and temperature time series (Table II). Additionally, mismatches on short timescales can propagate to longer timescales, as in the case of BCC-CSM1-1 [Fig. 3(1)]. Hence, explaining spectral properties of temperature time series not only requires consideration of stochastic noise,⁴¹ but also precise knowledge of its correlation structure and the forced response. This highlights a need for simple, stochastic dynamical models⁹⁹ to simulate temperature fluctuations on long timescales.

The components of AR5 EMICs show a reduced number of scales compared to the AOGCMs, which simplifies the complexity of the processes contributing to variability. Therefore, the relatively strong forced variability in EMICs [Figs. 4(a) and 4(b)] is not a main deficiency but serves to explore the long-term coupling between different Earth system components in response to radiative forcing. Compared to other AR5 EMICs, LOVECLIM V.1.2 features a more complex, three-layered atmosphere, which likely explains increased internal variations in the ensemble members. However, this variability is predominately short-term correlated. Due to the EBM's covariance structure, this can lead to an overestimation of the emulated forced+internal variability [Fig. 4(b)] on longer timescales. On interannual and decadal scales, the variance ratios based on the emulated forced and forced + internal variability from the LOVE-CLIM ensemble members [Fig. 4(b)] are similar to CMIP5 simulations [Fig. 4(c)]. The similarity indicates that EMICs with a more realistic representation of atmospheric variability might better capture the relative contribution of forced and internal temperature variations on these timescales.

The contributions of internal variations on multidecadal scales and longer remain the largest in CMIP5 models, likely due to longterm variability mechanisms from the comprehensive ocean dynamics of AOGCMs. Compared to observations [Fig. 2(d)], however, internal variability on decadal and multidecadal scales is smaller in CMIP5 simulations [Fig. 4(c)]. This is particularly interesting given the agreement of observed and simulated total GMST variability on these scales.^{2,3} Smaller low-frequency internal variability in climate models than in observations¹⁰⁰ could be offset by enhanced forced variability in response to volcanic eruptions,^{101,102} such that the overall variance is largely conserved. We suspect that an incorrect ratio between internal and external variability could impact the long-term variability of simulated local temperatures. However, uncertainties in the interpretation of our HadCRUT5 findings arise due to the comparatively short time-span of the instrumental record and a possible change of the forced response under global warming.¹⁰³ Further investigating the spatial variability structure¹⁰⁴ and the link between local and global variability across climate states could help resolve mismatches between observed and simulated local variability on decadal and multidecadal scales.

One limitation of our study arises from the fact that the developed framework targets interannual to centennial timescales and is, therefore, designed for annually resolved GMST and forcing data. As a result, it cannot be readily applied to much shorter or longer timescales. Investigating the immediate effects of radiative forcing, for example, necessitates an extension to sub-annual resolution, and treatment of the seasonal cycle. An extension to coarser resolutions could be beneficial to study long-term changes such as millennial-scale variability. However, such applications require careful examination of the underlying assumptions of the current framework, and an extension and validation of the estimation algorithm. The forced response is likely sensitive to model-specific rapid adjustments¹⁰⁵ and uncertainties in the forcing. Applying the workflow to paleoclimate reconstructions could, therefore, be challenging. On the one hand, "ClimBayes" does not yet run at the best possible speed, as there are faster Bayesian algorithms.²⁹ On the other hand, "ClimBayes" represents an accessible, transparent, and well-documented numerical framework that can be easily adapted and extended, for example, by integrating "Rstan"106 or multilevel delayed acceptance MCMC.¹⁰⁷ Similar to Fredriksen and Rypdal,¹⁵ we use a response function of exponential form, solving the ordinary differential equation (1). Future research could investigate the potential of Bayesian methods to find response functions describing the effects of climate forcing on different observables.^{108,109} Furthermore, future studies could test our findings with more advanced climate models including better representation of, for example, land surface processes, atmospheric dynamics and chemistry, and sea ice. The presented framework can be also applied to single forcing experiments⁷¹ for quantifying the contribution of single forcings to the spectrum. This will help better understand the climate system's response and interplay of intrinsic and external components in driving climate variability.

VI. CONCLUSION

We presented a physically motivated emulation of GMST data using Bayesian inference and a stochastic energy balance model. Analyzing AR5 EMICs and CMIP5 simulations for the last millennium, we found that the power spectral density of the combined *forced + internal* response approximates the *target* spectrum well. We show that our emulation can be used to separate internal and forced contributions to GMST variability across timescales. The relative contribution of internal dynamics increases with model complexity and decreases with timescale. While AR5 EMICs predominately exhibit forced variations, simulations from CMIP5 models and the LOVECLIM ensemble members exhibit major contributions from the forced and internal response. This suggests that

ARTICLE

EMICs with more realistic atmospheric variability can simulate statistical properties of interannual to decadal climate fluctuations more reliably. Our results show that precise knowledge of the forced response and correlation structure of internal variability is necessary to explain variability across scales, needed to assess future variability and potentially associated risks with long-term projections. Our developed framework is robust and readily available and can thus be widely applied to describe, emulate, and diagnose observed and simulated temperature variability.

ACKNOWLEDGMENTS

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, responsible for PMIP and CMIP. We thank the research groups listed in Table I and the Met Office for producing and making available their model output, measurements, and forcing reconstructions. We thank M. Eby for EMIC discussions, E. Myrvoll-Nilsen for discussion of Bayesian inference and the INLA package, and T. Gasenzer for discussion of conceptual climate models. We are grateful to N. Meinshausen, N. Weitzel, and E. Ziegler for helpful comments on the manuscript. We thank two anonymous referees for their constructive and valuable review. This study has been supported by funds of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 395588486, by the PalMod Project (Subproject No. 01LP1926C), the Heinrich-Böll-Stiftung (Heinrich Böll Foundation), and the Studienstiftung des deutschen Volkes (German Academic Scholarship Foundation). R. Scheichl is supported by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC No. 2181/1-390900948 (the Heidelberg STRUC-TURES Excellence Cluster). The study benefited from discussions within the CVAS working group, a working group of the Past Global Changes (PAGES) project.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

M.S. and B.E. contributed equally to this work. They carried out the analysis, created the figures and wrote the paper, supervised by K.R. and R.S. M.S. led the development of the ClimBayes package. B.E. led the variability analysis based on spectral methods. All authors designed the study, contributed to revisions, and approved the final version of the manuscript.

Maybritt Schillinger: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Resources (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Beatrice Ellerhoff:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Robert Scheichl:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal). **Kira Rehfeld:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The ClimBayes⁴² software package in R is released at https://github.com/paleovar/ClimBayes. and published under the Zenodo identifier https://doi.org/10.5281/zenodo.7317984. Code and data to reproduce all figures is available at https://github.com/pal eovar/EmulatingVariability. The data that support the findings of this study are openly available from the data holdings of the Climate Research Programme's Working Group on Coupled Modelling (e.g., https://esgf-node.llnl.gov/search/cmip5/), responsible for CMIP and PMIP, and from Schmidt *et al.*,⁴⁶ Eby *et al.*⁴⁵ (https://climate.uvic.ca/EMICAR5/participants.html), and Morice *et al.*⁴⁸

APPENDIX A: SOLUTION TO THE MULTIBOX EBM

The tridiagonal matrix K of the multibox EBM in matrix notation (3) is given by

$$\mathbf{K} = \begin{pmatrix} -(\tilde{\lambda} + \kappa_2) & \kappa_2 & 0 & \cdots & 0 \\ \kappa_2 & -(\kappa_2 + \kappa_3) & \kappa_3 & \ddots & \\ 0 & \kappa_3 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & & \\ 0 & \cdots & \kappa_N & -\kappa_N \end{pmatrix}.$$

The parameter λ controls the feedback of the surface layer. The coefficients $\kappa_2, \ldots, \kappa_N > 0$ describe the vertical heat transfer between ocean layers. The full solution to the multibox EBM¹⁵ reads

$$T = \int_{-\infty}^t e^{(t-s)C^{-1}K} C^{-1}F(s) \, ds.$$

K is symmetric and negative definite, and, thus, diagonalizable. Multiplication with the positive diagonal matrix C^{-1} does not change this property. Accordingly, the matrix exponential

$$e^{C^{-1}K} = V^T \begin{pmatrix} e^{-\lambda_1} & 0 & 0 & \cdots & 0\\ 0 & e^{-\lambda_2} & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & e^{-\lambda_N} \end{pmatrix} V$$

exists for an orthonormal matrix V and eigenvalues $-\lambda_k$ of $C^{-1}K$. Since only the first component of the forcing vector F(t) is non-zero, the matrix entry $(e^{C^{-1}K})_{11}$ defines the surface temperature response. The response function reads

$$R(t) = (e^{tC^{-1}K})_{11} = \sum_{k=1}^{N} w_k e^{-\lambda_k t}$$



FIG. 5. HadCRUT5 *target* observation, the posterior mean of the *forced* response of the stochastic two-box EBM, and three example realizations of the emulated *forced* + *internal* variability.

The normalization of the weights results from $\sum_{k=1}^{N} w_k$ = $\sum_{k=1}^{N} (V_{k1})^2 = 1$ for any orthonormal matrix *V*.

APPENDIX B: COVARIANCE OF THE ORNSTEIN-UHLENBECK AND AR(1) PROCESS

1. Covariance matrix

The noise term in Eq. (4) is a weighted sum of Ornstein-Uhlenbeck (OU) processes. Consequently, its covariance structure





results in

$$Cov(T_{1,I}(t), T_{1,I}(t+s)) = \frac{\sigma_W^2}{C_1^2} \sum_{k=1}^N \sum_{l=1}^N w_k w_l \frac{e^{-\lambda_l |s|}}{\lambda_k + \lambda_l}.$$
 (B1)

In the special case of N = 1, the covariance reduces to the covariance of a single OU process⁴⁰ Cov $(T_{1,I}(t), T_{1,I}(t+s))_{N=1} = \frac{\sigma_W^2}{2C_1^2\lambda_1} e^{-\lambda_1|s|}$. Formula (B1) follows from a generalization of this special case to arbitrary *N*. Discretizing the noise term in Eq. (4) results in a weighted sum of AR(1) processes. This sum *Z* is normally distributed with mean zero and covariance matrix

$$Cov(Z)_{ij} := Cov(Z_i, Z_j) = \frac{\sigma_W^2}{C_1^2} \sum_{k=1}^N \sum_{l=1}^N w_k w_l \frac{e^{-\lambda_l |i-j|}}{\lambda_k + \lambda_l}.$$
 (B2)

For given λ_k , w_k , and C_1 , the SD of the stochastic forcing, σ_W , uniquely defines the SD of the internal variations $\sigma_I = \sqrt{\text{Cov}(Z_i, Z_i)}$ and vice versa. It is possible to estimate σ_I within the Bayesian framework. However, additional uncertainties arise from the fact that the covariance matrix in Eq. (B2) is only an approximation to the true correlation structure of the residuals. As a result, the Bayesian estimation of σ_I might not preserve the total variance. Therefore, we determine σ_I from the residuals, that is, the data minus the estimated forced response.

Iterative computation of the likelihood

Theoretically, the likelihood is given by a normal distribution with mean $T_{1,F}$ and covariance matrix Cov(Z), depending on θ . However, computing the covariance matrix dynamically for each sample in the Markov chain can lead to difficulties. In particular, Cov(Z) needs to be inverted for every sample, which is computationally expensive. Moreover, the determinant det(Cov(Z)) can be close to zero, which can make numerical calculations unstable. Potential biases include decreasing goodness of fit and accuracy of estimated posteriors.

To solve this problem, we propose an iterative approach. This keeps λ_k and w_k in the covariance matrix fixed for each iteration of the algorithm. The first iteration uses the prior means for λ_k and w_k as well as a starting value for the ratio σ_W/C_1 . It is not necessary to consider σ_W and C_1 separately, since Eqs. (B1) and (B2) depend only on their ratio. This ratio is chosen such that $\sigma_I = 0.1$ K for CMIP5 simulations and LOVECLIM ensemble members, and $\sigma_I = 0.05$ K for AR5 EMICs. For the second iteration, the estimated posterior means of λ_k and w_k define the covariance matrix entries. Additionally, σ_I is set to the SD of the residuals, which defines σ_W/C_1 . The results of this second iteration are the posterior distributions for $\lambda_1, \ldots, \lambda_N$, weights w_2, \ldots, w_N , heat capacity C_1 , initial forcing F_0 , and initial temperature T_0 . These iterations can be repeated and adjusted with "ClimBayes." We find that two iterations are enough to fit the *forced* + *internal* response to the considered data well, and that further iterations do not improve the goodness of fit.

3. Sampling from internal variability

Sampling internal variations $T_{1,I}(t)$ requires the values of λ_k , w_k , and σ_W in Cov(Z) [Eq. (B1)]. λ_k and w_k are set to the posterior

Number of boxes	$\lambda_1 (yr^{-1})$	$\lambda_2 (yr^{-1})$	$\lambda_3 (yr^{-1})$	$C_1 (\mathrm{W}\mathrm{yr}\mathrm{m}^{-2}\mathrm{K}^{-1})$	w_1 (unitless)	w_2 (unitless)
1-box	0.35 (0.21,0.63)			9.08 (6.83,10.68)		
2-box	1.31 (0.71,1.86)	0.09 (0.03,0.16)		8.20 (5.68,10.43)	0.72 (0.46,0.9)	
3-box	1.29 (0.70,1.86)	0.11 (0.04,0.18)	0.01 (0.01,0.02)	8.46 (5.86,10.57)	0.74 (0.46,0.90)	0.22 (0.02,0.53)





FIG. 7. Emulated forced response $T_{1,F}(t)$, that is the posterior mean, of the stochastic two-box EBM fitted to the GMST target data from AR5 EMICs [panels (a)–(j)], LOVECLIM V.1.2 ensemble members [panels (k)–(0)], and CMIP5 simulations [panels (p)–(y)].

means. We calculate σ_W from σ_I , which we assume to equal the SD of the residuals. As an example, Fig. 5 shows realizations of $T_{1,I}(t)$ drawn from the estimated covariance of HadCRUT5 observations.

APPENDIX C: COMPARISON OF ONE-, TWO- AND THREE-BOX EBM

We have verified that our results are robust against reasonable variations of the number of boxes. Here, we examine the difference between $N \in (1, 2, 3)$ boxes on the example of the HadCRUT5 GMST (Fig. 6). We choose the priors of the response parameters to cover the same overall range as for the two-box model $[N = 1 : \lambda_1 \in (1/200, 2) \text{ yr}^{-1}$ and N = 3: $\lambda_1 \in (1/5, 2) \text{ yr}^{-1}, \lambda_2 \in (1/50, 1/5) \text{ yr}^{-1}, \lambda_3 \in (1/200, 1/50) \text{ yr}^{-1}].$

The stochastic two-box EBM fits the data more accurately (root mean square error: RMSE = 0.114 K) than the one-box EBM (RMSE = 0.133 K) (Fig. 6). The three-box EBM yields only minor improvements (RMSE = 0.0111 K). This pattern is consistent for AR5 EMICs and CMIP5 simulations and reflected in similar forced responses and power spectral densities for $N \in (1, 2, 3)$. Adding boxes, however, increases the risk of overfitting due to increasing degrees of freedom. This is reflected in increasing CIs for the forced response and parameters (Table III) with more boxes. That is why N = 2 represents the best compromise between goodness of fit, identifiability of parameters, and number of free parameters in our experiments.

APPENDIX D: EMULATED FORCED TEMPERATURE **RESPONSE FOR CONSIDERED SIMULATIONS**

Figure 7 shows the best estimate of the EBM's forced response, fitted to the target simulations from all considered models. CIs



FIG. 8. Ensemble mean of the CESM Large Ensemble¹¹⁰ (13 members) and the emulated forced response $T_{1,F}(t)$ of the stochastic two-box EBM fitted to GMST target data from one of the ensemble members (E1). The corresponding radiative forcing time series is shown in green.

are much narrower and almost vanishing compared to HadCRUT [Fig. 2(a)]. This is due to the fact that with increasing length of the time series the posterior uncertainties of the parameters and the forced response decrease.

Figure 8 compares the emulated forced response against simulation data from the Last Millennium Ensemble of the Community Earth System Model (CESM).¹¹⁰ Forming the ensemble mean over the available 13 members serves to average out uncorrelated internal variability. Our emulated forced response, fitted to the first ensemble member (E1), shows a large overlap with the ensemble mean despite remaining internal variability that has not been averaged out.

REFERENCES

¹R. W. Katz and B. G. Brown, "Extreme events in a changing climate: Variability is more important than averages," Clim. Change 21, 289-302 (1992).

²PAGES 2k Consortium, "Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era," Nat. Geosci. 12, 643-649 (2019).

³B. Ellerhoff and K. Rehfeld, "Probing the timescale dependency of local and global variations in surface air temperature from climate simulations and reconstructions of the last millennia," Phys. Rev. E 104, 064136 (2021).

⁴E. Hawkins and R. Sutton, "The potential to narrow uncertainty in projections of regional precipitation change," Clim. Dyn. 37, 407-418 (2011).

⁵L. M. Frankcombe, M. H. England, M. E. Mann, and B. A. Steinman, "Separating internal variability from the externally forced climate response," J. Clim. 28, 8184-8202 (2015).

⁶R. Hébert and S. Lovejoy, "Regional climate sensitivity- and historical-based projections to 2100," Geophys. Res. Lett. 45, 4248-4254, https://doi.org/10.1002/ 2017GL076649 (2018).

⁷T. Laepple and P. Huybers, "Global and regional variability in marine surface temperatures," Geophys. Res. Lett. 41, 2528-2534, https://doi.org/10.1002/2014G L059345 (2014).

⁸M. I. Budyko, "The effect of solar radiation variations on the climate of the earth," Tellus 21, 611-619 (1969).

⁹W. D. Sellers, "A global climatic model based on the energy balance of the earthatmosphere system," J. Appl. Meteorol. 8, 392-400 (1969).

10 K. Hasselmann, "Stochastic climate models: Part I. Theory," Tellus 28, 473-485 (1976).

11K. Fraedrich, U. Luksch, and R. Blender, "1/f model for long-time memory of the ocean surface temperature," Phys. Rev. E 70, 037301 (2004).

¹²I. M. Held, M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, "Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing," J. Clim. 23, 2418-2427 (2010).

13 O. Geoffroy, D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, and S. Tytéca, "Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments," J. Clim. 26, 1841–1857 (2013). ¹⁴M. Rypdal and K. Rypdal, "Long-memory effects in linear response models

of earth's temperature and implications for future global warming," J. Clim. 27, 5240-5258 (2014).

¹⁵H.-B. Fredriksen and M. Rypdal, "Long-range persistence in global surface temperatures explained by linear multibox energy balance models," J. Clim. 30, 7157-7168 (2017). ¹⁶K. Hasselmann, "Optimal fingerprints for the detection of time-dependent

climate change," J. Clim. 6, 1957-1971 (1993).

¹⁷G. C. Hegerl, H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, "Detecting greenhouse-gas-induced climate change with an optimal fingerprint method," J. Clim. 9, 2281-2306 (1996).

¹⁸M. Ghil, "Climate sensitivity, energy balance models, and oscillatory climate models," J. Geophys. Res. 89, 1280, https://doi.org/10.1029/JD089iD01p01280 (1984).

¹⁹P. M. Cox, C. Huntingford, and M. S. Williamson, "Emergent constraint on equilibrium climate sensitivity from global temperature variability," Nature 553, 319–322 (2018).

²⁰K. Rypdal, M. Rypdal, and H.-B. Fredriksen, "Spatiotemporal long-range persistence in earth's temperature field: Analysis of stochastic–diffusive energy balance models," J. Clim. **28**, 8379–8395 (2015).

²¹M. Ghil, "Climate stability for a sellers-type model," J. Atmos. Sci. 33, 3-20 (1976).

²²G. R. North, J. G. Mengel, and D. A. Short, "Simple energy balance model resolving the seasons and the continents: Application to the astronomical theory of the ice ages," J. Geophys. Res. **88**, 6576, https://doi.org/10.1029/JC088iC11p06576 (1983).

²³T. Bódai, V. Lucarini, F. Lunkeit, and R. Boschi, "Global instability in the Ghil-Sellers model," Clim. Dyn. 44, 3361–3381 (2015).

²⁴L. Østvand, T. Nilsen, K. Rypdal, D. Divine, and M. Rypdal, "Long-range memory in internal and forced dynamics of millennium-long climate model simulations," Earth Syst. Dyn. 5, 295–308 (2014).

²⁵G. R. North, "Analytical solution to a simple climate model with diffusive heat transport," J. Atmos. Sci. **32**, 1301–1307 (1975).

²⁶G. R. North and R. F. Cahalan, "Predictability in a solvable stochastic climate model," J. Atmos. Sci. 38, 504–513 (1981).

²⁷G. R. North, J. Wang, and M. G. Genton, "Correlation models for temperature fields," J. Clim. **24**, 5850–5862 (2011).

 ²⁸B. Dortmans, W. F. Langford, and A. R. Willms, "An energy balance model for paleoclimate transitions," Clim. Past 15, 493–520 (2019).
 ²⁹E. Myrvoll-Nilsen, S. H. Sørbye, H.-B. Fredriksen, H. Rue, and M. Rypdal, "Sta-

²⁹E. Myrvoll-Nilsen, S. H. Sørbye, H.-B. Fredriksen, H. Rue, and M. Rypdal, "Statistical estimation of global surface temperature response to forcing under the assumption of temporal scaling," Earth Syst. Dyn. **11**, 329–345 (2020).

³⁰R. Hébert, S. Lovejoy, and B. Tremblay, "An observation-based scaling model for climate sensitivity estimates and global projections to 2100," Clim. Dyn. 56, 1105–1129 (2021).

³¹R. W. Bodman and R. N. Jones, "Bayesian estimation of climate sensitivity using observationally constrained simple climate models," WIREs Clim. Change 7, 461–473 (2016).

 ⁵²C. Proistosescu and P. J. Huybers, "Slow climate mode reconciles historical and model-based estimates of climate sensitivity," Sci. Adv. 3, e1602821 (2017).
 ³³A. Jonko, N. M. Urban, and B. Nadiga, "Towards Bayesian hierarchical infer-

⁵³A. Jonko, N. M. Urban, and B. Nadiga, "Towards Bayesian hierarchical inference of equilibrium climate sensitivity from a combination of CMIP5 climate models and observational data," Clim. Change **149**, 247–260 (2018).

³⁴S. C. Sherwood, M. J. Webb, J. D. Annan, K. C. Armour, P. M. Forster, J. C. Hargreaves, G. Hegerl, S. A. Klein, K. D. Marvel, E. J. Rohling, M. Watanabe, T. Andrews, P. Braconnot, C. S. Bretherton, G. L. Foster, Z. Hausfather, A. S. Heydt, R. Knutti, T. Mauritsen, J. R. Norris, C. Proistosescu, M. Rugenstein, G. A. Schmidt, K. B. Tokarska, and M. D. Zelinka, "An assessment of earth's climate sensitivity using multiple lines of evidence," Rev. Geophys. 58, e2019RG000678, https://doi.org/10.1029/2019RG000678 (2020).

³⁵R. B. Skeie, T. Berntsen, M. Aldrin, M. Holden, and G. Myhre, "Climate sensitivity estimates—Sensitivity to radiative forcing time series and observational data," Earth Syst. Dyn. **9**, 879–894 (2018).

³⁶J. M. Mitchell, "An overview of climatic variability and its causal mechanisms," Quat. Res. 6, 481–493 (1976).

³⁷P. Huybers and W. Curry, "Links between annual, Milankovitch and continuum temperature variability," Nature **441**, 329–332 (2006).

³⁸A. S. von der Heydt, P. Ashwin, C. D. Camp, M. Crucifix, H. A. Dijkstra, P. Ditlevsen, and T. M. Lenton, "Quantification and interpretation of the climate variability record," Global Planet. Change **197**, 103399 (2021).

³⁹C. L. E. Franzke, S. Barbosa, R. Blender, H.-B. Fredriksen, T. Laepple, F. Lambert, T. Nilsen, K. Rypdal, M. Rypdal, M. G. Scotto, S. Vannitsem, N. W. Watkins, L. Yang, and N. Yuan, "The structure of climate variability across scales," Rev. Geophys. 58, e2019RG000657, https://doi.org/10.1029/2019RG000657 (2020).

⁴⁰M. Rypdal, H.-B. Fredriksen, E. Myrvoll-Nilsen, K. Rypdal, and S. Sørbye, "Emergent scale invariance and climate sensitivity," Climate 6, 93 (2018).

⁴¹S. A. Soldatenko and R. A. Colman, "Power spectrum sensitivity analysis of the global mean surface temperature fluctuations simulated in a two-box stochastic energy balance model," Tellus A: Dyn. Meteorol. Oceanogr. **74**, 68 (2022).

⁴²M. Schillinger, B. Ellerhoff, R. Scheichl, and K. Rehfeld, "The ClimBayes package in R," Zenodo, V. 0.1.1, https://doi.org/10.5281/zenodo.7317984 (2022).
⁴³J. H. Jungclaus, E. Bard, M. Baroni, P. Braconnot, J. Cao, L. P. Chini, T. Egorova, M. Evans, J. F. González-Rouco, H. Goosse, G. C. Hurtt, F. Joos, J. O. Kaplan, M. Khodri, K. Klein Goldewijk, N. Krivova, A. N. LeGrande, S. J. Lorenz, J. Luterbacher, W. Man, A. C. Maycock, M. Meinshausen, A. Moberg, R. Muscheler, C. Nehrbass-Ahles, B. I. Otto-Bliesner, S. J. Phipps, J. Pongratz, E. Rozanov, G. A. Schmidt, H. Schmidt, W. Schmutz, A. Schurer, A. I. Shapiro, M. Sigl, J. E. Smerdon, S. K. Solanki, C. Timmreck, M. Toohey, I. G. Usoskin, S. Wagner, C.-J. Wu, K. L. Yeo, D. Zanchettin, Q. Zhang, and E. Zorita, "The PMIP4 contribution to CMIP6—Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 *past1000* simulations," Geosci. Model Dev. **10**, 4005–4033 (2017).

⁴⁴G. Flato, J. Marotzke, B. Abiodun, P. Braconnot, S. C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, and C. Forest, "Evaluation of climate models," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2014), pp. 741–866.

⁴⁵M. Eby, A. J. Weaver, K. Alexander, K. Zickfeld, A. Abe-Ouchi, A. A. Cimatoribus, E. Crespin, S. S. Drijfhout, N. R. Edwards, A. V. Eliseev, G. Feulner, T. Fichefet, C. E. Forest, H. Goosse, P. B. Holden, F. Joos, M. Kawamiya, D. Kicklighter, H. Kienert, K. Matsumoto, I. I. Mokhov, E. Monier, S. M. Olsen, J. O. P. Pedersen, M. Perrette, G. Philippon-Berthier, A. Ridgwell, A. Schlosser, T. Schneider von Deimling, G. Shaffer, R. S. Smith, R. Spahni, A. P. Sokolov, M. Steinacher, K. Tachiiri, K. Tokos, M. Yoshimori, N. Zeng, and F. Zhao, "Historical and idealized climate model experiments: An intercomparison of earth system models of intermediate complexity," Clim. Past 9, 1111–1140 (2013).

⁴⁶G. A. Schmidt, J. H. Jungclaus, C. M. Ammann, E. Bard, P. Braconnot, T. J. Crowley, G. Delaygue, F. Joos, N. A. Krivova, R. Muscheler, B. L. Otto-Bliesner, J. Pongratz, D. T. Shindell, S. K. Solanki, F. Steinhilber, and L. E. A. Vieira, "Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1)," Geosci. Model Dev. 5, 185–191 (2012).

⁴⁷K. Tachiiri, J. C. Hargreaves, J. D. Annan, A. Oka, A. Abe-Ouchi, and M. Kawamiya, "Development of a system emulating the global carbon cycle in earth system models," Geosci. Model Dev. 3, 365–376 (2010).
⁴⁸C. P. Morice, J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J.

**C. P. Morice, J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J. H. Dunn, T. J. Osborn, P. D. Jones, and I. R. Simpson, "An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set," J. Geophys. Res.: Atmos. 126, e2019JD032361, https://doi.org/10.1029/2019JD032361 (2021).

⁴⁹J. Pongratz, C. Reick, T. Raddatz, and M. Claussen, "A reconstruction of global agricultural areas and land cover for the last millennium," <u>Global Biogeochem</u>. <u>Cycles</u> **22**, GB3018, https://doi.org/10.1029/2007GB003153 (2008).

⁵⁰T. J. Crowley, G. Zielinski, B. Vinther, R. Udisti, K. Kreutz, J. Cole-Dai, and E. Castellano, "Volcanism and the little ice age," PAGES News 16, 22–23 (2008).
 ⁵¹F. Steinhilber, J. Beer, and C. Fröhlich, "Total solar irradiance during the

⁵¹F. Steinhilber, J. Beer, and C. Fröhlich, "Total solar irradiance during the holocene," Geophys. Res. Lett. **36**, L19704, https://doi.org/10.1029/2009GL040142 (2009).

²²Y.-M. Wang, J. L. Lean, and J. N. R. Sheeley, "Modeling the sun's magnetic field and irradiance since 1713," Astrophys. J. **625**, 522–538 (2005).

⁵³G. Delaygue and E. Bard, "An Antarctic view of beryllium-10 and solar activity for the past millennium," Climate Dyn. **36**(11–12), 2201–2218 (2011).

⁵⁴N. A. Krivova, L. Balmaceda, and S. K. Solanki, "Reconstruction of solar total irradiance since 1700 from the surface magnetic flux," Astron. Astrophys. 467, 335–346 (2007).

⁵⁵L. E. A. Vieira and S. K. Solanki, "Evolution of the solar magnetic flux on time scales of years to millenia," Astron. Astrophys. **509**, A100 (2010).

⁵⁶C. Gao, A. Robock, and C. Ammann, "Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models," J. Geophys. Res. **113**, D23111, https://doi.org/10.1029/2008JD010239 (2008).

⁵⁷S. P. Ritz, T. F. Stocker, and F. Joos, "A coupled dynamical ocean-energy balance atmosphere model for paleoclimate studies," J. Clim. 24, 349–375 (2011).
⁵⁸M. Montoya, A. Griesel, A. Levermann, J. Mignot, M. Hofmann, A. Ganopolski, and S. Rahmstorf, "The earth system model of intermediate complexity CLIMBER-3α. Part I: Description and performance for present-day conditions," Clim. Dyn. 26, 327–328 (2006).

⁵⁹V. Petoukhov, M. Claussen, A. Berger, M. Crucifix, M. Eby, A. V. Eliseev, T. Fichefet, A. Ganopolski, H. Goosse, I. Kamenkovich, I. I. Mokhov, M. Montoya, L. A. Mysak, A. Sokolov, P. Stone, Z. Wang, and A. J. Weaver, "EMIC intercomparison project (EMIP-CO₂): Comparative analysis of EMIC simulations of climate, and of equilibrium and transient responses to atmospheric CO₂ doubling," Clim. Dyn. 25, 363–385 (2005).

⁶⁰G. Shaffer, S. Malskær Olsen, and J. O. Pepke Pedersen, "Presentation, calibration and validation of the low-order, DCESS earth system model (Version 1)," Geosci. Model Dev. 1, 17–51 (2008).

⁶¹ A. P. Sokolov, C. A. Schlosser, S. Dutkiewicz, S. Paltsev, D. W. Kicklighter, H. D. Jacoby, R. G. Prinn, C. E. Forest, J. M. Reilly, C. Wang, and B. S. Felzer, "MIT integrated global system model (IGSM) version 2: Model description and baseline evaluation," Tech. Rep. (MIT Joint Program on the Science and Policy of Global Change, 2005).

⁶²H. Goosse, V. Brovkin, T. Fichefet, R. Haarsma, P. Huybrechts, J. Jongma, A. Mouchet, F. Selten, P.-Y. Barriat, J.-M. Campin, E. Deleersnijder, E. Driesschaert, H. Goelzer, I. Janssens, M.-F. Loutre, M. A. Morales Maqueda, T. Opsteegh, P.-P. Mathieu, G. Munhoven, E. J. Pettersson, H. Renssen, D. M. Roche, M. Schaeffer, B. Tartinville, A. Timmermann, and S. L. Weber, "Description of the earth system model of intermediate complexity LOVECLIM version 1.2," Geosci. Model Dev. 3, 603–633 (2010).

⁶³K. Matsumoto, K. S. Tokos, A. R. Price, and S. J. Cox, "First description of the minnesota earth system model for ocean biogeochemistry (MESMO 1.0)," Geosci. Model Dev. 1, 1–15 (2008).

⁶⁴N. Zeng, "How strong is carbon cycle-climate feedback under global warming?," Geophys. Res. Lett. **31**, L20203, https://doi.org/10.1029/2004GL020904 (2004).

⁶⁵A. J. Weaver, M. Eby, E. C. Wiebe, C. M. Bitz, P. B. Duffy, T. L. Ewen, A. F. Fanning, M. M. Holland, A. MacFadyen, H. D. Matthews, K. J. Meissner, O. Saenko, A. Schmittner, H. Wang, and M. Yoshimori, "The UVic earth system climate model: Model description, climatology, and applications to past, present and future climates," Atmos. Ocean **39**, 361–428 (2001).

⁶⁶X. Xiao-Ge, W. Tong-Wen, and Z. Jie, "Introduction of CMIP5 experiments carried out with the climate system models of Beijing climate center," Adv. Clim. Change Res. **4**, 41–49 (2013).

⁶⁷L. Landrum, B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom, and H. Teng, "Last millennium climate and its variability in CCSM4," J. Clim. **26**, 1085–1111 (2013).

⁶⁸S. J. Phipps, L. D. Rotstayn, H. B. Gordon, J. L. Roberts, A. C. Hirst, and W. F. Budd, "The CSIRO Mk3L climate system model version 1.0—Part 2: Response to external forcings," Geosci. Model Dev. 5, 649–682 (2012).

⁶⁹Q. Bao, P. Lin, T. Zhou, Y. Liu, Y. Yu, G. Wu, B. He, J. He, L. Li, J. Li, Y. Li, H. Liu, F. Qiao, Z. Song, B. Wang, J. Wang, P. Wang, X. Wang, Z. Wang, B. Wu, T. Wu, Y. Xu, H. Yu, W. Zhao, W. Zheng, and L. Zhou, "The flexible global ocean-atmosphere-land system model, spectral version 2: FGOALS-s2," Adv. Atmos. Sci. **30**, 561–576 (2013).

⁷⁰G. A. Schmidt, R. Ruedy, J. E. Hansen, I. Aleinov, N. Bell, M. Bauer, S. Bauer, B. Cairns, V. Canuto, Y. Cheng, A. Del Genio, G. Faluvegi, A. D. Friend, T. M. Hall, Y. Hu, M. Kelley, N. Y. Kiang, D. Koch, A. A. Lacis, J. Lerner, K. K. Lo, R. L. Miller, L. Nazarenko, V. Oinas, J. Perlwitz, J. Perlwitz, D. Rind, A. Romanou, G. L. Russell, M. Sato, D. T. Shindell, P. H. Stone, S. Sun, N. Tausnev, D. Thresher, and M.-S. Yao, "Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data," J. Clim. **19**, 153–192 (2006).

⁷¹ A. P. Schurer, S. F. B. Tett, and G. C. Hegerl, "Small influence of solar variability on climate over the past millennium," Nat. Geosci. 7, 104–108 (2014).

⁷²C. D. Jones, J. K. Hughes, N. Bellouin, S. C. Hardiman, G. S. Jones, J. Knight, S. Liddicoat, F. M. O'Connor, R. J. Andres, C. Bell, K.-O. Boo, A. Bozzo, N. Butchart, P. Cadule, K. D. Corbin, M. Doutriaux-Boucher, P. Friedlingstein, J. Gornall, L. Gray, P. R. Halloran, G. Hurtt, W. J. Ingram, J.-F. Lamarque, R. M. Law, M. Meinshausen, S. Osprey, E. J. Palin, L. Parsons Chini, T. Raddatz, M. G. Sanderson, A. A. Sellar, A. Schurer, P. Valdes, N. Wood, S. Woodward, M. Yoshioka, and M. Zerroukat, "The HadGEM2-ES implementation of CMIP5 centennial simulations," Geosci. Model Dev. 4, 543–570 (2011).

⁷³J.-L. Dufresne, M.-A. Foujols, S. Denvil, A. Caubel, O. Marti, O. Aumont, Y. Balkanski, S. Bekki, H. Bellenger, R. Benshila, S. Bony, L. Bopp, P. Braconnot, P. Brockmann, P. Cadule, F. Cheruy, F. Codron, A. Cozic, D. Cugnet, N. de Noblet, J.-P. Duvel, C. Ethé, L. Fairhead, T. Fichefet, S. Flavoni, P. Friedlingstein, J.-Y.

Grandpeix, L. Guez, E. Guilyardi, D. Hauglustaine, F. Hourdin, A. Idelkadi, J. Ghattas, S. Joussaume, M. Kageyama, G. Krinner, S. Labetoulle, A. Lahellec, M.-P. Lefebvre, F. Lefevre, C. Levy, Z. X. Li, J. Lloyd, F. Lott, G. Madec, M. Mancip, M. Marchand, S. Masson, Y. Meurdesoif, J. Mignot, I. Musat, S. Parouty, J. Polcher, C. Rio, M. Schulz, D. Swingedouw, S. Szopa, C. Talandier, P. Terray, N. Viovy, and N. Vuichard, "Climate change projections using the IPSL-CM5 earth system model: From CMIP3 to CMIP5," Clim. Dyn. 40, 2123–2165 (2013).

⁷⁴F. Hourdin, M.-A. Foujols, F. Codron, V. Guemas, J.-L. Dufresne, S. Bony, S. Denvil, L. Guez, F. Lott, J. Ghattas, P. Braconnot, O. Marti, Y. Meurdesoif, and L. Bopp, "Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model," Clim. Dyn. **40**, 2167–2192 (2013).

⁷⁵T. Sueyoshi, R. Ohgaito, A. Yamamoto, M. O. Chikamoto, T. Hajima, H. Okajima, M. Yoshimori, M. Abe, R. O'ishi, F. Saito, S. Watanabe, M. Kawamiya, and A. Abe-Ouchi, "Set-up of the PMIP3 paleoclimate experiments conducted using an earth system model, MIROC-ESM," Geosci. Model Dev. 6, 819–836 (2013).

⁷⁶M. A. Giorgetta, J. Jungclaus, C. H. Reick, S. Legutke, J. Bader, M. Böttinger, V. Brovkin, T. Crueger, M. Esch, K. Fieg, K. Glushak, V. Gayler, H. Haak, H.-D. Hollweg, T. Ilyina, S. Kinne, L. Kornblueh, D. Matei, T. Mauritsen, U. Mikolajewicz, W. Mueller, D. Notz, F. Pithan, T. Raddatz, S. Rast, R. Redler, E. Roeckner, H. Schmidt, R. Schnur, J. Segschneider, K. D. Six, M. Stockhause, C. Timmreck, J. Wegner, H. Widmann, K.-H. Wieners, M. Claussen, J. Marotzke, and B. Stevens, "Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5: Climate changes in MPI-ESM," J. Adv. Model. Earth Syst. 5, 572–597 (2013).

⁷⁷J. H. Jungclaus, N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, and J. S. Storch, "Characteristics of the ocean simulations in the max planck institute ocean model (MPIOM) the ocean component of the MPI-earth system model," J. Adv. Model. Earth Syst. **5**, 422–446 (2013).

⁷⁸D. P. Cummins, D. B. Stephenson, and P. A. Stott, "Optimal estimation of stochastic energy balance model parameters," J. Clim. 33, 7909–7926 (2020).

⁷⁹O. Geoffroy, D. Saint-Martin, G. Bellon, A. Voldoire, D. J. L. Olivié, and S. Tytéca, "Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs," J. Clim. **26**, 1859–1876 (2013).

⁸⁰ A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," Stat. Sci. 7, 457–472 (1992).

⁸¹ S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," J. Comput. Graph. Stat. 7, 434–455 (1998).

⁸²A. Gelman, "Bayesian data analysis," in *Chapman & Hall/CRC Texts in Statistical Science*, 3rd ed. (CRC Press, Boca Raton, FL, 2014).

⁸³M. Plummer, N. Best, K. Cowles, K. Vines, and M. M. Plummer, "R package 'CODA'" https://cran.r-project.org/package=coda (2020).

⁸⁴J. M. Flegal, M. Haran, and G. L. Jones, "Markov chain Monte Carlo: Can we trust the third significant figure?," Stat. Sci. 23, 250–260 (2008).

⁸⁵M. Haran and J. Hughes, "R package 'batchmeans: Consistent batch means estimation of Monte Carlo," https://cran.r-project.org/package=batchmeans (2020).

⁸⁶T. Laepple and P. Huybers, "Ocean surface temperature variability: Large model data differences at decadal and longer periods," Proc. Natl. Acad. Sci. U.S.A. 111, 16682–16687 (2014).

⁸⁷O. Bothe, J. H. Jungclaus, and D. Zanchettin, "Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble," Clim. Past **9**, 2471–2487 (2013).

⁸⁸R. Zhang, T. L. Delworth, R. Sutton, D. L. R. Hodson, K. W. Dixon, I. M. Held, Y. Kushnir, J. Marshall, Y. Ming, R. Msadek, J. Robson, A. J. Rosati, M. Ting, and G. A. Vecchi, "Have aerosols caused the observed atlantic multidecadal variability?," J. Atmos. Sci. **70**, 1135–1144 (2013).

⁸⁹D. Olonscheck and D. Notz, "Consistently estimating internal climate variability from climate model simulations," J. Clim. **30**, 9555–9573 (2017).

⁹⁰F. Lehner, C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti, and E. Hawkins, "Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6," Earth Syst. Dyn. **11**, 491–508 (2020).

⁹¹ M. E. Mann, B. A. Steinman, D. J. Brouillette, A. Fernandez, and S. K. Miller, "On the estimation of internal climate variability during the preindustrial past millennium," Geophys. Res. Lett. **49**, e2021GL096596, https://doi.org/10.1029/2021GL096596 (2022).

⁹²M. E. Mann, B. A. Steinman, and S. K. Miller, "On forced temperature changes, internal variability, and the AMO," Geophys. Res. Lett. **41**, 3211–3219, https://doi.org/10.1029/JD089iD01p01280 (2014).

⁹³T. J. Crowley, "Causes of climate change over the past 1000 years," Science 289, 270–277 (2000).

⁹⁴J. E. Kay, C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. M. Arblaster, S. C. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner, J.-F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein, "The community earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability," Bull. Am. Meteorol. Soc. **96**, 1333–1349 (2015).

(2015). ⁹⁵D. G. MacMynowski, H.-J. Shin, and K. Caldeira, "The frequency response of temperature and precipitation in a climate model," Geophys. Res. Lett. **38**, L16711, https://doi.org/10.1029/2011GL048623 (2011).

⁹⁶N. Maher, D. Matei, S. Milinski, and J. Marotzke, "ENSO change in climate projections: Forced response or internal variability?," Geophys. Res. Lett. 45, 11390–11398, https://doi.org/10.1029/2018GL079764 (2018).

⁹⁷R. Bonnet, D. Świngedouw, G. Gastineau, O. Boucher, J. Deshayes, F. Hourdin, J. Mignot, J. Servonnat, and A. Sima, "Increased risk of near term global warming due to a recent AMOC weakening," Nat. Commun. **12**, 6108 (2021).

⁹⁸B. Ellerhoff, M. J. Kirschner, E. Ziegler, M. D. Holloway, L. Sime, and K. Rehfeld, "Contrasting state-dependent effects of natural forcing on global and local climate variability," Geophys. Res. Lett. **49**, e2022GL098335, https://doi.org/10.1029/2022GL098335 (2022).

⁹⁹S. Lovejoy, "The half-order energy balance equation—Part 1: The homogeneous HEBE and long memories," Earth Syst. Dyn. **12**, 469–487 (2021).

¹⁰⁰X. Yan, R. Zhang, and T. R. Knutson, "Underestimated AMOC variability and implications for AMV and predictability in CMIP models," Geophys. Res. Lett. 45, 4319–4328, https://doi.org/10.1029/2018GL077378 (2018). ¹⁰¹ A. P. Schurer, G. C. Hegerl, M. E. Mann, S. F. B. Tett, and S. J. Phipps, "Separating forced from chaotic climate variability over the past millennium," J. Clim. 26, 6954–6973 (2013).
¹⁰² P. Chylek, C. Folland, J. D. Klett, and M. K. Dubey, "CMIP5 climate

¹⁰² P. Chylek, C. Folland, J. D. Klett, and M. K. Dubey, "CMIP5 climate models overestimate cooling by volcanic aerosols," Geophys. Res. Lett. 47, e2020GL087047, https://doi.org/10.1029/2020GL087047 (2020).
 ¹⁰³ P. O. Hopcroft, J. Kandlbauer, P. J. Valdes, and R. S. J. Sparks, "Reduced

¹⁰³P. O. Hopcroft, J. Kandlbauer, P. J. Valdes, and R. S. J. Sparks, "Reduced cooling following future volcanic eruptions," Clim. Dyn. **51**, 1449–1463 (2018).

¹⁰⁴T. Kunz and T. Laepple, "Frequency-dependent estimation of effective spatial degrees of freedom," J. Clim. **34**, 7373–7388 (2021).

¹⁰⁵R. S. Smith, J. M. Gregory, and A. Osprey, "A description of the FAMOUS (version XDBUA) climate model and control run," Geosci. Model Dev. 1, 53–68 (2008).

[106] Guo, D. Lee, K. Sakrejda, J. Gabry, B. Goodrich, J. De Guzman, E. Niebler, T. Heller, and J. Fletcher, "Rstan: R interface to stan," https://cran.r-project.org/package=rstan (2022).

¹⁰⁷M. B. Lykkegaard, T. J. Dodwell, C. Fox, G. Mingas, and R. Scheichl, "Multilevel delayed acceptance MCMC," arXiv:2202.03876 (2022).

¹⁰⁸V. Lucarini, "Revising and extending the linear response theory for statistical mechanical systems: Evaluating observables as predictors and predictands," J. Stat. Phys. **173**, 1698–1721 (2018).

¹⁰⁹G. L. Torres Mendonça, J. Pongratz, and C. H. Reick, "Identification of linear response functions from arbitrary perturbation experiments in the presence of noise—Part 1: Method development and toy model demonstration," Nonlinear Processes Geophys. 28, 501–532 (2021).

¹¹⁰ B. L. Otto-Biesner, E. C. Brady, J. Fasullo, A. Jahn, L. Landrum, S. Stevenson, N. Rosenbloom, A. Mai, and G. Strand, "Climate variability and change since 850 CE: An ensemble approach with the community earth system model," Bull. Am. Meteorol. Soc. **97**, 735–754 (2016).