Ocean surface temperature variability: Large model–data differences at decadal and longer periods

Thomas Laepple[a,1] and Peter Huybers[b]

[a]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, 13353 Potsdam, Germany; and [b]Department of Earth and Planetary Science, Harvard University, Cambridge, MA 02138

Edited by Mark H. Thiemens, University of California, San Diego, La Jolla, CA, and approved September 23, 2014 (received for review July 14, 2014)

The variability of sea surface temperatures (SSTs) at multidecadal and longer timescales is poorly constrained, primarily because instrumental records are short and proxy records are noisy. Through applying a new noise filtering technique to a global network of late Holocene SST proxies, we estimate SST variability between annual and millennial timescales. Filtered estimates of SST variability obtained from coral, foraminifer, and alkenone records are shown to be consistent with one another and with instrumental records in the frequency bands at which they overlap. General circulation models, however, simulate SST variability that is systematically smaller than instrumental and proxy-based estimates. Discrepancies in variability are largest at low latitudes and increase with timescale, reaching two orders of magnitude for tropical variability at millennial timescales. This result implies major deficiencies in observational estimates or model simulations, or both, and has implications for the attribution of past variations and prediction of future change.

sea surface temperature | climate variability | multiproxy synthesis | proxy data reconstruction

Variations in sea surface temperature (SSTs) have widespread implications for society and are the basis of most regional decadal prediction efforts (1). Magnitudes of variability in regional SSTs are inferred either using observations or simulations from general circulation models (GCMs). At synoptic and interannual timescales, there is overall agreement between observational and GCM estimates of SST variability (2–4). At decadal timescales, however, instrumental records generally show greater regional SST variability than found in the Coupled Model Intercomparison Project Phase 5 (CMIP5) ensemble of GCM simulations (4) and in earlier simulations (5–7). Discrepancies are greatest at low latitudes where model–data mismatches in variance reach a factor of 2 (Fig. 1).

Estimating regional SST variability at multidecadal and longer timescales presents a quandary. Discrepancies with instrumentally observed SST variability at decadal timescales calls into question the credibility of GCM estimates at longer timescales. At the same time, instrumental observations covering more than 100 y of SST variability are sparse. For example, whereas 68% of ocean grid boxes in the Climate Research Unit's instrumental compilation of SSTs have a 30-y interval with an observational density of at least 100 observations per year, only 19% of grid boxes have such coverage over a 100-y interval (8). Furthermore, SST variability observed during the last century represents contributions from natural and anthropogenic sources that are difficult to disentangle and are not necessarily representative of any other interval (9).

One is inevitably led to using paleoclimate proxies to constrain multidecadal and longer-term variability. Numerous paleoclimate reconstructions document low-frequency SST variability (10–15), but the degree to which these reconstructions afford quantitative constraints of annual average SST variability is unclear because of issues including noise, sampling artifacts, and possible proxy-specific biases. In the following, we apply a recently introduced procedure for filtering out artifacts from proxy spectral estimates (16) to a global dataset of high-resolution proxies to estimate SST variability from interannual to millennial timescales.

We next systematically compare these observational estimates against results from long GCM simulations. Although foregoing studies have analyzed proxy–model agreement for global average temperature (17) and other climate indices (18), to our knowledge, this is the first proxy–model comparison of regional SST variability.

Proxy Data

Building on an existing compilation (19), we analyze 33 high-resolution proxy records of marine temperature sampled from marine sediment cores (Uk37 and planktonic Mg/Ca) and corals (Sr/Ca, $\delta^{18}O$, and growth rate) (Fig. 2). We use paleo observations extending back no more than 7,000 y because observations from earlier in the Holocene may be influenced by transient adjustments associated with the last deglaciation. Proxy results are compared against instrumental SSTs (8) and long GCM simulations (20–23) (*SI Appendix*, Tables S1 and S2). To facilitate intercomparison, all proxy records are recalibrated using a single temperature relationship for each proxy type (*SI Appendix*, *Proxy Calibration*). The importance of using a consistent calibration can be seen in that the spatial correlation between instrumental and coral-derived interannual temperature variance is 0.84, whereas it would only be 0.26 were individually published calibrations instead used (*SI Appendix*, Fig. S4).

The spectra of each proxy record in our collection is estimated using a multitaper procedure. Results indicate that centennial and millennial variations are substantially larger than those found at decadal timescales (Fig. 3*A*), as has been previously noted (10, 11, 14, 24), but the magnitude and shape of the spectral estimates

Significance

Determining magnitudes of sea surface temperature variability is important for attributing past and predicting future changes in climate, and generally requires the use of proxies to constrain multidecadal and longer timescales of variability. We report a multiproxy estimate of sea surface temperature variability that is consistent between proxy types and with instrumental estimates but strongly diverges from climate model simulations toward longer timescales. At millennial timescales, model–data discrepancies reach two orders of magnitude in the tropics, indicating substantial problems with models or proxies, or both, and highlighting a need to better determine the variability of sea surface temperatures.

Author contributions: T.L. and P.H. designed research; T.L. and P.H. performed research; T.L. analyzed data; and T.L. and P.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: tlaepple@awi.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1412077111/-/DCSupplemental.

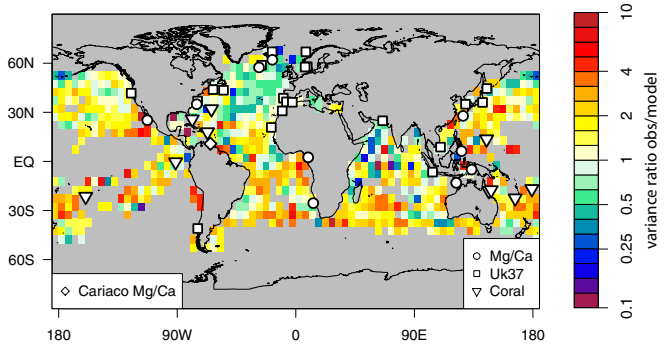

Fig. 1. Variance ratio of the observed and simulated sea surface temperature variance for multidecadal (20–50 y) timescales from ref. 4. Simulated variance is from the mean variance of all CMIP5 historical simulations. Observed variance is from HadSST3 (8) and is corrected for sampling and instrumental errors. Locations of the temperature proxies used in this study are marked as symbols.

between different proxy types is incommensurate. At overlapping frequencies, the Mg/Ca records average 2.1 times more variance than the Uk37 records and proportionately more high- than low-frequency variability. Mg/Ca estimates also show an order of magnitude more variability than indicated by coral records where their resolved frequencies overlap, near 1/200 y^{-1}. These discrepancies must be resolved if proxies are to provide a plausible estimate of temperature variability or afford a credible test of GCM simulations of SST variability (25).

One possibility is that regional differences in temperature variability account for the discrepancies between proxies. Nearby Mg/Ca and Uk37 records are, however, in no better agreement. Furthermore, the core site locations associated with Mg/Ca measurements do not show any greater variability than the Uk37 sites in either instrumental records or GCM simulations (16). Discrepancies between coral and Mg/Ca or alkenone estimates are similarly unresolved by spatial variability. Instead, it appears that differences in variability between proxy types arise from differential noise influences.

Noise sources can be grouped into three categories involving errors associated with measurement noise, mixing of sediments by biological activity that leads to smoothing of measured signals, called bioturbation (26), and aliasing of high-frequency variability associated with irregular or infrequent sampling (27, 28). Other sources of uncertainty are also present, but we will show that the above well-recognized noise sources are sufficient to resolve interproxy discrepancies. Our approach is to design a filter for each proxy record to remove the noise effects (or amplify variability where it has been suppressed by bioturbation) depending on proxy type, sediment accumulation rates, and local SST variability. See *Materials and Methods* and ref. 16. for details.

There are four indications that the proxy correction technique gives accurate results. First, the technique recovers true SST variability in expectation when applied to synthetic records. Second, the inferred noise values correspond with independently derived error estimates available for four of the Mg/Ca records (16). Third, the corrected proxy spectral estimates for corals, Mg/Ca, and Uk37 are consistent among

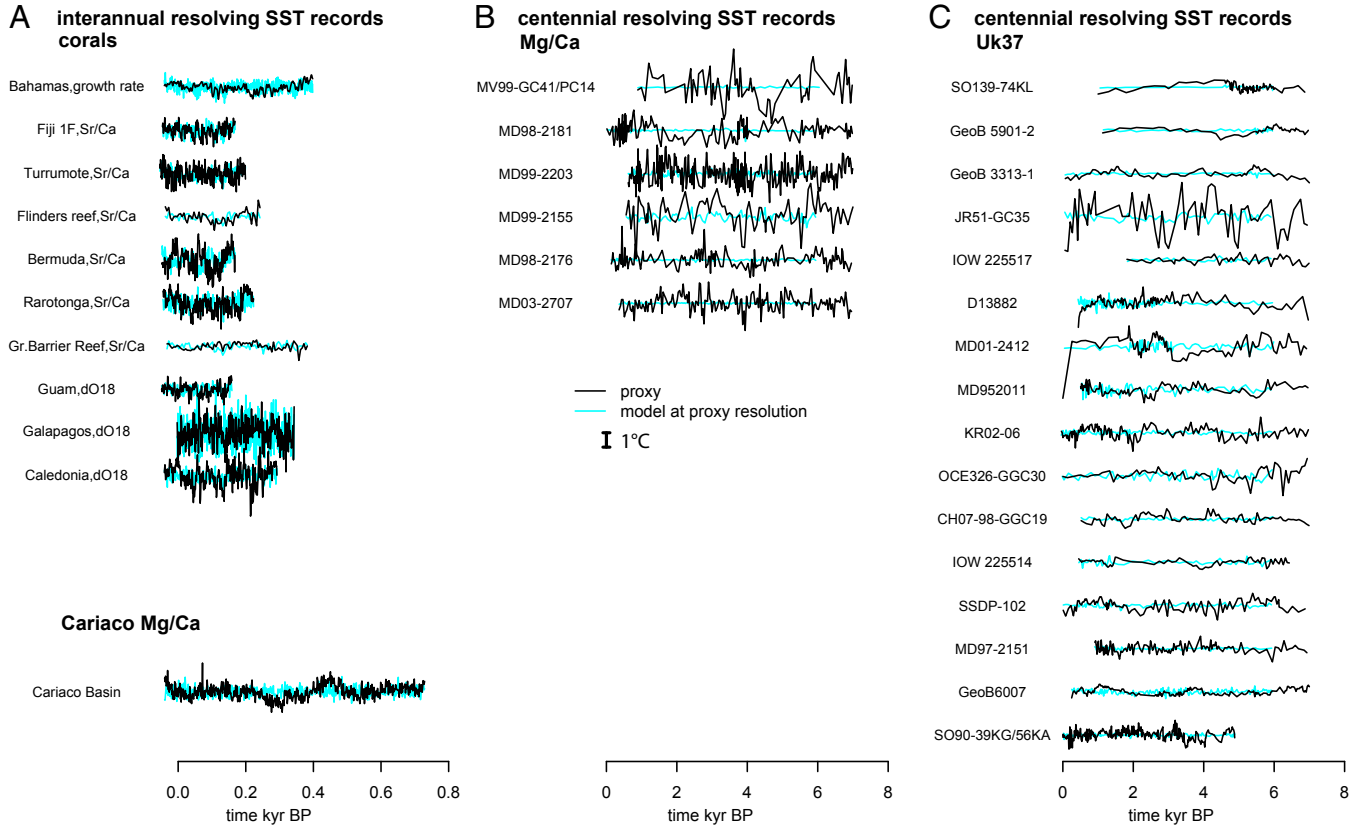

Fig. 2. Raw proxy data (black) plotted against the corresponding model simulations sampled at the proxy locations (cyan). Both simulations and proxies are detrended, and the simulations are temporally averaged to the same resolution as the proxies. To provide continuous temporal overlap with the proxy records, interannual resolution records (*A*) are compared against a comprehensively forced simulation (20), whereas centennial resolution records (*B* and *C*) are compared against a GCM simulation forced only by orbital variations (21). Differences in variability between model simulations are small relative to model–data discrepancies.

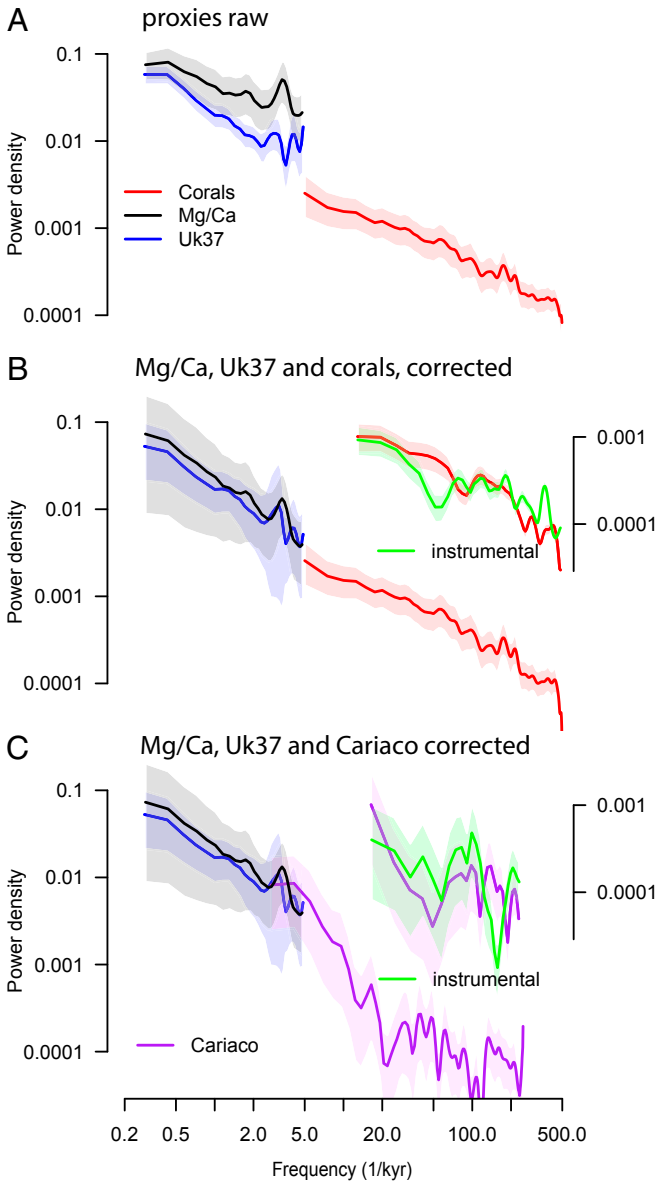

Fig. 3. Spectral estimates from proxy and instrumental SSTs. (*A*) Raw proxy spectral estimates. (*B*) Proxy spectral estimates after correction for measurement noise, bioturbational smoothing, and aliasing. (*C*) Same as *B* but showing the Cariaco Mg/Ca based SST spectra instead of corals. The Cariaco Mg/Ca record is not included in *A* and *B*. *Insets* show high-resolution proxy spectra and HadSST3 instrumental spectra estimated using the time interval in which they overlap. All other spectra are calculated over their full duration for years before 1950. The 95% confidence intervals are indicated by shading, and account for uncertainties in the Uk37 and Mg/Ca correction process where appropriate.

themselves in terms of magnitude and relative proportions of variability (Fig. 3 *B* and *C*). Greater reduction in Mg/Ca proxy variance (−53%) relative to that of Uk37 (−25%) and corals (−12%) results from the small numbers of individual samples combined together for each Mg/Ca estimate and the larger aliasing correction that is consequently required. Fourth and finally, records having a frequency resolution that overlaps with the instrumental record—corals and the high-resolution Mg/Ca record from the Cariaco Basin (10)—are consistent to within uncertainties with spectral estimates from local instrumental SST records (Fig. 3 *B* and *C*, *Insets*).

Model Simulations

Numerous long GCM simulations exist that can be evaluated against our proxy results (*SI Appendix*, Table S2). We focus on a five-member ensemble of simulations from the ECHAM5/MPIOM Earth system model that extends over the last 1,200 y and includes forcing associated with variations in solar output, Earth's orbital configuration, atmospheric greenhouse gas concentrations, volcanic aerosols, and land use changes (20). We also analyze longer simulations from the same model with no forcing (20) or only using orbital forcing (21). All ECHAM5/MPIOM simulations produce similar local SST variability (Fig. 4) and have essentially constant spectral energy at frequencies below 1/50 y^{-1}.

In addition to ECHAM/MPIOM, we also analyze the 15 long-term historical simulations from across the 8 models contained in the CMIP5/Paleoclimate Modelling Intercomparison Project Phase III (PMIP3) last millenium ensemble (22, 23) (*SI Appendix*, Figs. S2 and S3). The ECHAM5/MPIOM simulations produce SST variability in the upper range of the CMIP5/PMIP3 simulations across all timescales. It follows that model–data differences discussed with respect to ECHAM5/MPIOM, hereafter simply referred to as the model, also hold more generally.

Results

Proxy spectral estimates (Fig. 3) show substantially greater variability than found in model simulations (Fig. 4), with the discrepancy increasing from interannual to millennial timescales (Fig. 5). There is also a clear meridional structure in the model–data mismatch that persists across timescales (Fig. 5). The tropically amplified pattern of model–data mismatch found at decadal timescales using instrumental records (Fig. 1) is also found at longer timescales (Fig. 5), but with proxy estimates showing systematically greater variability. At low latitudes (30°S–30°N), the proxy estimates average 45 times greater variance than is associated with model SST variability [95% confidence interval (c.i.) of 22–92] at multicentennial timescales (Fig. 5*D*), and this discrepancy grows to more than a factor of 100 (c.i. 118–1,470) at millennial timescales (Fig. 5*E*). Although averages can be sensitive to outliers, similar order-of-magnitude discrepancies are found when medians are instead used, giving

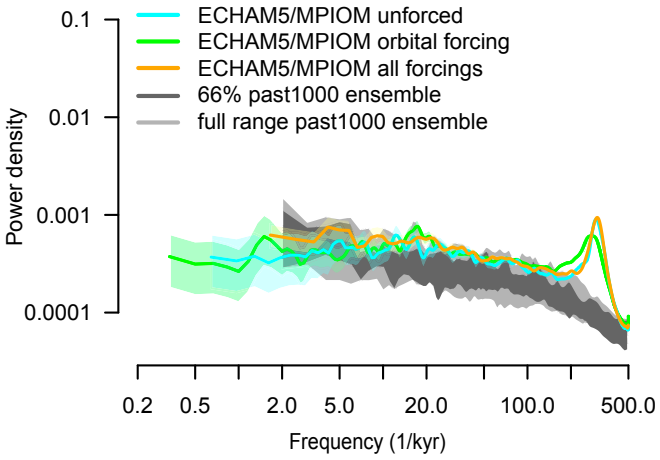

Fig. 4. SST spectral estimates from ECHAM5/MPIOM and the CMIP5/PMIP3 collection of simulations. The ECHAM5/MPIOM spectral estimates are from simulations that are unforced, forced only by changes in Earth's orbital configuration, and forced using orbital, solar, volcanic, greenhouse gas, and land-used changes. The range of spectral estimates from the fully forced CMIP5/PMIP3 past1000 simulations are also shown. In all cases, spectral energy is nearly constant at frequencies below 1/50 y^{-1}.

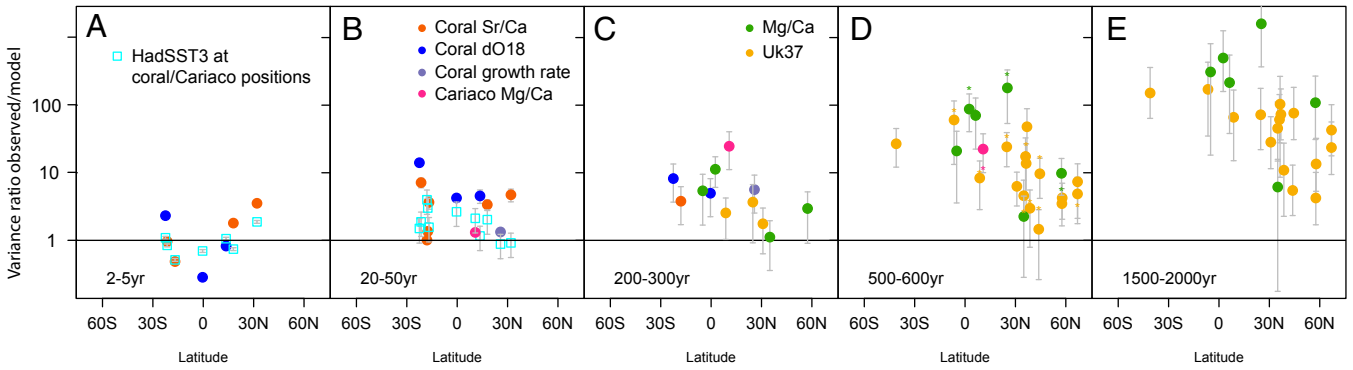

Fig. 5. Latitudinal dependence of the model–data mismatch at different timescales. Shown is the variance ratio between observed and GCM simulated sea surface temperature, where simulations are sampled at the data positions and observational estimates are corrected for sampling and measurement errors. Proxy comparisons (dots) are relative to different versions of the ECHAM5/MPIOM model results: at timescales less than 1,000 y (*A*–*D*), comparisons are from the fully forced results, whereas comparisons at millennial timescales (*E*) are for the orbital-only results. Ratios derived from the orbital-only simulation are also shown as small stars in *D* and demonstrate that results are consistent using either model version. Variance ratios from instrumental observations (HADSST3) at the proxy positions (cyan square) are shown in *A* and *B*. Error bars show 67% confidence intervals, comparable to a SE, and include uncertainties associated with the correction process (see *Materials and Methods*).

ratios of 23 (c.i. 8–41) and 120 (c.i. 15–290) at multicentennial and millennial timescales, respectively.

Discrepancies between model and data SST spectral estimates are smaller at midlatitudes to high latitudes (>30°N or S). At decadal timescales (Fig. 1), the models have slightly greater variance than the instrumental observations. At multicentennial timescales (Fig. 5*D*), proxy estimates average 10 (c.i. 6–18) times greater variance, and at millennial timescales, 49 (c.i. 34–151) times greater variance (Fig. 5*E*). Medians give 6 times (c.i. 3–8) and 35 times (c.i. 13–50) the variance at multicentennial and millennial timescales, respectively. Essentially the same results are obtained from either instrumental or coral records at high frequencies, from either Mg/Ca or Uk37 records at low frequencies, and using any of the CMIP5/PMIP3 simulations (*SI Appendix*, Figs. S2 and S3), indicating that the present results are robust and generalizable.

In considering possible explanations for model–data discrepancies, we begin with the data. One possibility is that model grid boxes are unrepresentative of the scale of variability sampled by the proxies, but this seems unlikely because instrumental and proxy records agree with the models at interannual timescales (Figs. 3 and 5 *A* and *B*). Furthermore, the spatial coherence of temperature variability generally increases toward longer timescales (9) and discrepancies between point and grid box estimates would be expected to correspondingly diminish toward longer timescales, if gridding were at issue. Another issue is that proxy sites are disproportionately located in regions near the ocean margins—where higher sedimentation rates afford higher-resolution records. However, we find no systematic difference in model–data discrepancy between coastal and more interior sites, in agreement with other findings that coastal point observations and open ocean gridded SSTs have similar variability (29).

Another data issue is for foraminifera and alkenone-producing species to record specific seasons or dwell in subsurface regions having different temperature variability, but sampling the variability within single seasons and at different depth levels fails to resolve the model–data mismatch (*SI Appendix*, Fig. S6). Temporal changes in growth seasonality or depth habitat could lead to spurious proxy variability, and alkenone records may also be subject to time variable resuspension and transport (30). These processes are not identifiable in our results, however, in that estimates agree across the three different classes of proxies (Fig. 3), similar latitudinal model–data discrepancies are found when considering either the instrumental or proxy records (Fig. 5), and similar scaling is found in Mg/Ca records from distinct foram types (i.e., ruber and bulloides). Moreover, studies of foraminifera generally indicate that, in seeking homeostasis, there is a tendency toward muting, as opposed to increasing, the magnitude of recorded temperature variability (31). There are also sources of uncertainty associated with calibrating coral proxy records of sea surface temperature variability (32), but comparison between different coral-based SST proxy types and between coral and instrumental data suggests no systematic bias (*SI Appendix*, *Reliability of Coral-Based Estimates of Marine Variability*, and Figs. S4 and S5).

A final data issue is that we may have systematically underestimated noise contributions. An independent check of noise variance is available through analyzing the covariance of nearby records, and we find that these are consistent with the noise levels inferred from our proxy-correction approach (*SI Appendix*, *Signal-to-Noise Estimates*, and Fig. S1). Under the assumption that noise is uncorrelated between records, estimated signal-to-noise ratios are an order of magnitude too great for noise to explain the model–data mismatch. Although the net effect of the noise correction that we apply is to decrease proxy and instrumental variance, we find no evidence that proxy variance should be further reduced.

Discussion

The combined instrumental and proxy evidence indicates that models systematically underestimate regional temperature variability and that this mismatch increases toward longer timescales. This result is consistent with land surface temperatures reconstructed from tree rings, other terrestrial proxies, and documentary evidence also indicating greater regional variability than simulated by models at decadal and longer timescales (33–35). That the ECHAM5/MPIOM model simulation of the carbon cycle over the last millenium fails to reproduce the magnitude of preindustrial atmospheric CO_2 variability (20) may also result from too little natural variability. We also note that models generally fail to simulate the magnitude of climate variations indicated by more ancient paleoclimate records (36).

One possible reconciliation for model–data discrepancies in variability is for the models to have insufficient internal variability. For instance, ocean–atmosphere coupling may be too weak (37), or the energy cascade from the mesoscale to larger spatial scales (38) may be insufficient. These scenarios could also be related to models being too diffusive, as would be consistent with model–data discrepancies growing toward longer timescales and the fact that higher-resolution simulations in the CMIP5 ensemble tend to show greater regional variability (4).

Another means of model–data reconciliation is through greater natural external forcing. For example, regional SST variance is roughly doubled at multidecadal and centennial timescales in the GISS-E2-R ensemble of simulations (39) when using larger-magnitude volcanic forcing estimates (40) relative to smaller ones (41) (*SI Appendix*, Fig. S7). As another example, regional SST variability is twice as large at centennial and millennial timescales during early Holocene portions of the TraCE-21ka experiment (42) when ice melt contributes significant amounts of freshwater, compared with late Holocene variability when ice melt is set to zero (*SI Appendix*, Fig. S7). These results suggest that increased variability in freshwater fluxes, perhaps associated with shifts in rainfall, would also increase regional SST variability. A related set of possibilities involves underestimating the sensitivity to external forcing. We are not aware of transient simulations that produce variability as large as that indicated by the proxy record, including those that represent volcanic (39), freshwater (42), or solar forcing (43). Notably, however, ref. 43 demonstrates a similar pattern of response to solar variability in both proxy data and simulations, and that an equilibrium simulation of the Maunder Minimum (44) having a detailed representation of the stratosphere showed strong regional temperature responses to solar forcing, similar in magnitude to those reconstructed elsewhere from proxies. Together these examples illustrate that additional forcing and amplification mechanisms readily increase regional SST variability and may possibly resolve model–data discrepancies.

Of course, it is also possible to generate consistent magnitudes of variability for the wrong reasons. For example, consistent variability was found between a simulation and paleoclimate reconstruction of the El Niño Southern Oscillation, but where the simulation showed greater volcanic contributions than found in the reconstruction (18). More detailed study of forcing-response patterns seems important for identifying plausible mechanisms by which to reconcile model–proxy discrepancies. An obstacle, however, is that whereas age-model errors have little influence on the spectral estimation relied upon in this study (45), such age-model errors generally corrupt covariance estimates. Accurate analysis of forcing-response patterns using proxy data will likely require development of new statistical approaches that better account for timing errors.

Conclusions

Given the large number of parameterized contributions that combine to determine regional SST variability, it is not surprising that models do not reproduce magnitudes of variability out to arbitrarily long timescales. Similarly, the large number of possible influences upon the proxy record makes it difficult to rule out systematic biases in variance. Nonetheless, obtaining accurate estimates and simulations of regional SST variability is important in several respects. From a historical perspective, an accurate reconstruction of the magnitude of past changes is obviously preferable. There are also implications for attribution and prediction of changes. Testing whether changes in temperature are attributable to anthropogenic causes requires a null distribution representing natural variability that is typically obtained from control runs of a model (5). Insomuch as models underestimate natural SST variability, tests for anthropogenic effects will tend to be biased positive. Optimal detection techniques that seek to rotate the fingerprint of climate change so as to maximize signal-to-noise ratios in model simulations are likely even more susceptible to bias (46). Underestimation of internal variability can also be expected to lead to predicted ranges that are too narrow, possibly also because of the projection of anthropogenic forcing onto natural modes of variability (1).

Reconciliation of model–data mismatch in SST variability requires participation from the modeling, statistics, and paleoclimate communities. Continued study of the magnitude and sensitivity to forcing, model representation of SST variability, and the processes through which proxies record environmental change all appear important. It may also be relevant to seek a more complete representation of marine proxies within general circulation models, including the growth and life cycle of marine proxies (31) as well as the processes through which paleoclimate signals are recorded and preserved in sediments.

Materials and Methods

Dataset Selection and Sampling. Only mid to late Holocene proxy records with a mean resolution of less than 100 y and a length of more than 4,000 y are included in the analysis. Coral records were chosen on the basis of being at least 200 y long and reported to mainly record SST. The Cariaco Mg/Ca record is the only annually resolved sediment record considered in this study. We limit our collection to records having relatively high resolution to facilitate skillful correction for the effects of measurement noise, bioturbation, and sampling (16). All time series are evenly interpolated before analysis. To minimize aliasing, data are first linearly interpolated to 10 times the target resolution, low-pass filtered using a finite response filter with a cutoff frequency of 1.2 divided by the target time step, and then resampled at the target resolution. Note that linear interpolation tends to reduce the variance near the Nyquist frequencies of a process that has been unevenly sampled (45), and we have therefore excluded variance estimates at the highest frequencies. The appropriate resolution for each proxy record is determined by sampling power-law processes according to the native resolution of a given proxy and empirically determining which frequencies are unbiased—usually those below about four times the mean sampling frequency (*SI Appendix*, Table S1).

Proxy Correction. Mg/Ca records are subject to nonnegligible levels of errors from all three error categories (measurement noise, bioturbation, and aliasing of high-frequency variability associated with irregular or infrequent sampling), and we correct for these through an extension of the filtering approach of Kirchner (27). Measurement noise here includes instrumental error and vital effects associated with foraminifera that are related to biologically controlled influences on the calcification process. Our approach is to design synthetic temperature records whose spectra are consistent with those from the observations once the appropriate sources of noise have been applied. Synthetic Mg/Ca records are sampled according to seasonal growth times (31), bioturbated (26) (for unlaminated records), and subsampled according to the number of individuals combined into each actual proxy measurement. Each Mg/Ca subsample is then corrupted by independent realizations of normally distributed noise. Synthetic Uk37 records are similarly bioturbated and subjected to sampling noise, but the effects of aliasing are not included because Uk37 samples comprise millions of molecules from across many different years. A filter is estimated for each Mg/Ca or Uk37 record by taking the ratio of the spectra associated with the synthetic record before and after subjecting it to the various sources of error. Application of this filter to the actual record then yields our best estimate of the actual SST variability. See ref. 16 for a complete description, detailed evaluation, and application of this technique to centennially resolved Mg/Ca and Uk37 records.

Correcting the annually resolved coral, Cariaco (10), and instrumental records is simpler because bioturbation and aliasing can be ignored, and we simply subtract the noise contribution following the errors reported in the original publication (*SI Appendix*, Table S1). Except for the case of correcting for bioturbation, all adjustments lead to a decrease in the SST variance inferred from each indicator. The net result of applying our correction algorithms is to decrease the average variance associated with each proxy type as well as that of the instrumental record.

Spectral Estimation. Spectra are estimated using Thomson's multitaper method (47) with three windows. Time series are detrended before analysis, as it is standard for spectral estimation. The multitaper approach introduces a small bias at the lowest frequencies, and we omit the two lowest frequencies in Figs. 3 and 4. For visual display purposes, power spectra are smoothed using a Gaussian kernel with constant width in logarithmic frequency space (27) when using logarithmic axes. To avoid biased estimates at low- and high-frequency boundaries, the kernel is truncated on both sides to maintain its symmetry. Except for instrumental comparisons (Figs. 3 *B* and *C*, *Insets*, and Fig. 5 *A* and *B*), years after 1950 are not spectrally analyzed, to omit an interval that likely contains substantial nonstationarity.

The average power spectrum for each proxy type shown in Fig. 3 necessarily contains samples from regions with differing variability and that cover different frequency intervals. To avoid discontinuities across frequencies

where the number of available estimates change, proxy spectra are scaled to an average value in the largest common frequency interval. Error bars in Fig. 3 account for this weighting. Note that ratio calculations (Fig. 5) neither require nor contain scaling. Mean regional SST spectral estimates (Fig. 4 and *SI Appendix*, Fig. S7) are derived by calculating the spectra for every ocean grid box between 50°S and 50°N and showing the area-weighted mean of all regional spectra. Confidence intervals for spectral estimates are obtained from Monte Carlo simulations, and they account for the effects associated with corrections to the proxy records, the weighting associated with combining different spectral estimates, the frequency-dependent smoothing kernel, and the normal uncertainties associated with a spectral estimate. Specifically, we simulate surrogate time series using the estimated β scaling relationship (16), then corrupt the records, correct them, and estimate the resulting spectra. The process is repeated one thousand times, and a χ^2 distribution is fit to the ensemble of results at each frequency using moment matching. Uncertainties in ratio estimates are estimated analogously except that they are approximated as following an F distribution. For purposes of averaging, individual proxy spectra are assumed to be independent. Likewise, each of the five ensemble members from the fully forced millennial GCM simulations are assumed independent.

Ratio of Proxy and GCM SST Variance. Timescale-dependent variance ratios (Fig. 5) are derived by summing spectral energy estimates between respective frequencies. The confidence intervals for the mean and median ratios are derived by drawing realizations of ratios for every proxy from an F distribution, taking the mean or median, repeating this 100,000 times, and reporting the quantiles.

Comparison With Instrumental SST. Instrumental SST variance estimates that are used for purposes of comparison come from the corrected HadSST3 variance spectra from ref. 16. Comparisons are made with respect to the grid box containing a given proxy record. We do not restrict the instrumental SST data to have at least 10 observations per year because this restriction would leave large data gaps in the instrumental records at the coral positions. For comparing the coral and Cariaco records against instrumental SSTs (Fig. 3 *B* and *C*, *Insets*), only overlapping years are analyzed for proxy and instrumental data. For the ratio between observed and model SST variability (Fig. 5), all years for which model results and each type of data overlap are included.

ACKNOWLEDGMENTS. We thank Johann Jungclaus and Nils Fischer for providing the ECHAM5/MPIOM-ESM simulations; Tom Marchitto, Peter deMenocal, Caroline Cleroux, Peter Stott, Eva Calvo, and Elizabeth Farmer for information regarding replicates; Braddock Linsley, Jung-Hyun Kim, and Guillaume Leduc for providing proxy data; and Stephen Barker, Thomas Felis, John Kennedy, Gerrit Lohmann, Cristian Proistosescu, Andy Rhines, Dan Schrag, Stephan Dietrich, Tilla Roy, and Alexander Stine for helpful discussion. We acknowledge the World Climate Research Programmes Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. T.L. was supported by the Alexander von Humboldt Foundation, the Daimler and Benz Foundation, and the Initiative and Networking Fund of the Helmholtz Association Grant VG-NH900. P.H. received support from NSF Awards 0960787 and 1060454.

1. Solomon A, et al. (2011) Distinguishing the roles of natural and anthropogenically forced decadal climate variability: Implications for prediction. *Bull Am Meteorol Soc* 92:141–156.
2. Collins M, Tett SFB, Cooper C (2001) The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Clim Dyn* 17(1):61–81.
3. Min SK, Legutke S, Hense A, Kwon WT (2005) Internal variability in a 1000-yr control simulation with the coupled climate model ECHO-G – I. Near-surface temperature, precipitation and mean sea level pressure. *Tellus Series A* 57(4):605–621.
4. Laepple T, Huybers P (2014) Global and regional variability in marine surface temperatures. *Geophys Res Lett* 41:2528–2534.
5. Stott PA, Tett SFB (1998) Scale-dependent detection of climate change. *J Clim* 11: 3282–3294.
6. Davey M, et al. (2002) STOIC: A study of coupled model climatology and variability in tropical ocean regions. *Clim Dyn* 18:403–420.
7. DelSole T (2006) Low-frequency variations of surface temperature in observations and simulations. *J Clim* 19:4487–4507.
8. Kennedy JJ, Rayner NA, Smith RO, Parker DE, Saunby M (2011) Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J Geophys Res* 116(D14):D14103.
9. Jones PD, Osborn TJ, Briffa KR (1997) Estimating sampling errors in large-scale temperature averages. *J Clim* 10:2548–2568.
10. Black DE, et al. (2007) An 8-century tropical Atlantic SST record from the Cariaco Basin: Baseline variability, twentieth-century warming, and Atlantic hurricane frequency. *Paleoceanography* 22(4):PA4204.
11. deMenocal P, Ortiz J, Guilderson T, Sarnthein M (2000) Coherent high- and low-latitude climate variability during the Holocene warm period. *Science* 288(5474):2198–2202.
12. Cobb KM, Charles CD, Cheng H, Edwards RL (2003) El Niño/Southern Oscillation and tropical Pacific climate during the last millennium. *Nature* 424(6946):271–276.
13. Kim J, et al. (2007) Impacts of the North Atlantic gyre circulation on Holocene climate off northwest Africa. *Geology* 35:387–390.
14. Sachs J (2007) Cooling of Northwest Atlantic slope waters during the Holocene. *Geophys Res Lett* 34(3):L03609.
15. Farmer EJ, Chapman MR, Andrews JE (2008) Centennial-scale Holocene North Atlantic surface temperatures from Mg/Ca ratios in Globigerina bulloides. *Geochem Geophys Geosyst* 9(12):Q12029.
16. Laepple T, Huybers P (2013) Reconciling discrepancies between UK37 and Mg/Ca reconstructions of Holocene marine temperature variability. *Earth Planet Sci Lett* 375: 418–429.
17. Crowley TJ (2000) Causes of climate change over the past 1000 years. *Science* 289(5477):270–277.
18. Ault T, Deser C, Newman M, Emile-Geay J (2013) Characterizing decadal to centennial variability in the equatorial Pacific during the last millennium. *Geophys Res Lett* 40: 3450–3456.
19. Leduc G, Schneider R, Kim JH, Lohmann G (2010) Holocene and Eemian sea surface temperature trends as revealed by alkenone and Mg/Ca paleothermometry. *Quat Sci Rev* 29:989–1004.
20. Jungclaus JH, et al. (2010) Climate and carbon-cycle variability over the last millennium. *Clim. Past* 6:723–737.
21. Fischer N, Jungclaus J (2011) Evolution of the seasonal temperature cycle in a transient Holocene simulation: Orbital forcing and sea-ice. *Clim Past* 7:1139–1148.
22. Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498.
23. Braconnot P, et al. (2012) Evaluation of climate models using palaeoclimatic data. *Nat Clim Change* 2:417–424.
24. Huybers P, Curry W (2006) Links between annual, Milankovitch and continuum temperature variability. *Nature* 441(7091):329–332.
25. Barnett T, et al. (1999) Detection and attribution of recent climate change: A status report. *Bull Am Meteorol Soc* 80:26312660.
26. Berger W, Heath G (1968) Vertical mixing in pelagic sediments. *J Mar Res* 26:134143.
27. Kirchner JW (2005) Aliasing in 1/*f*($^{\alpha}$) noise spectra: Origins, consequences, and remedies. *Phys Rev E Stat Nonlin Soft Matter Phys* 71(6 Pt 2):066110.
28. Laepple T, Werner M, Lohmann G (2011) Synchronicity of Antarctic temperatures and local solar insolation on orbital timescales. *Nature* 471(7336):91–94.
29. MacKenzie BR, Schiedek D (2007) Long-term sea surface temperature baselines time series, spatial covariation and implications for biological processes. *J Mar Syst* 68: 405–420.
30. Ohkouchi N, Eglinton TI, Keigwin LD, Hayes JM (2002) Spatial and temporal offsets between proxy records in a sediment drift. *Science* 298(5596):1224–1227.
31. Fraile I, et al. (2009) Modeling the seasonal distribution of planktonic foraminifera during the Last Glacial Maximum. *Paleoceanography* 24:PA2216.
32. Scott RB, Holland CL, Quinn TM (2010) Multidecadal trends in instrumental SST and coral proxy Sr/Ca records. *J Clim* 23:1017–1033.
33. Collins M, Osborn TJ, Tett SF, Briffa KR, Schweingruber FH (2002) A comparison of the variability of a climate model with paleotemperature estimates from a network of tree-ring densities. *J Clim* 15:1497–1515.
34. Goosse H, Renssen H, Timmermann A, Bradley RS (2005) Internal and forced climate variability during the last millennium: A model-data comparison using ensemble simulations. *Quat Sci Rev* 24:1345–1360.
35. Zorita E, et al. (2010) European temperature records of the past five centuries based on documentary/instrumental information compared to climate simulations. *Clim Change* 101:143–168.
36. Valdes P (2011) Built for stability. *Nat Geosci* 4:414–416.
37. Cane MA (1998) A role for the tropical Pacific. *Science* 282:59–61.
38. Ferrari R, Wunsch C (2009) Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annu Rev Fluid Mech* 41:253–282.
39. Schmidt GA, et al. (2014) Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J Adv Model Earth Syst* 6:141–184.
40. Gao C, Robock A, Ammann C (2008) Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *J Geophys Res* 113(D23): D23111.
41. Crowley TJ, et al. (2008) Volcanism and the little ice age. *PAGES News* 16:22–23.
42. Liu Z, et al. (2009) Transient simulation of last deglaciation with a new mechanism for Bolling-Allerod warming. *Science* 325(5938):310–314.
43. Moffa-Sánchez P, Born A, Hall IR, Thornalley DJR, Barker S (2014) Solar forcing of North Atlantic surface temperature and salinity over the past millennium. *Nat Geosci* 7:275–278.
44. Shindell DT, Schmidt GA, Mann ME, Rind D, Waple A (2001) Solar forcing of regional climate change during the Maunder Minimum. *Science* 294(5549):2149–2152.
45. Rhines A, Huybers P (2011) Estimation of spectral power laws in time uncertain series of data with application to the Greenland Ice Sheet Project 2 δ^{18}O record. *J Geophys Res* 116(D1):D01103.
46. Allen MR, Tett SF (1999) Checking for model consistency in optimal fingerprinting. *Clim Dyn* 15:419–434.
47. Percival DB, Walden AT (1993) *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques* (Cambridge Univ Press, Cambridge, UK).

Supplementary information for: How variable are marine sea surface temperatures beyond decadal time scales?

Thomas Laepple * **and Peter Huybers** †

*Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany, and †Department of Earth and Planetary Science, Harvard University, Cambridge, MA 02138, USA

Proceedings of the National Academy of Sciences of the United States of America

SI-1. Overview

The supplementary information is divided into the description of the proxy calibration (SI-2), a discussion of coral based variance estimates (SI-3) an independent check of the signal/noise variance in the proxies (SI-4) and sensitivity tests of our results relative to choices in habitat depth, seasonality, bioturbation depth and our proxy-record selection (SI-5).

SI-2. Proxy calibration

All proxy records of a given type are uniformly calibrated to facilitate spatial comparison (Fig. 5). $U_{37}^{k'}$ records are calibrated using $0.033U_{37}^{K'}/^{\circ}C$; U_{37}^{k} records using $0.035U_{37}^{K'}/^{\circ}C$; and Mg/Ca records, including the Cariaco Basin record [1], using 9.35% Mg/Ca per °C. These choices are the mean of all author calibrations of the analysed datasets and are similar to other standard calibrations: $0.033U_{37}^{K'}/^{\circ}C$ [2] and 9% Mg/Ca per °C [3]. The single exception is a growth rate coral record [4] for which no independent calibration exists. For Sr/Ca and δ^{18}O in corals there is evidence that calibrations based on seasonal relationships lead to an overestimation of interannual and longer temperature variability, possibly because interannual signals are attenuated by calcification occurring at some depth into the coral [5]. We therefore recalibrate all coral estimates to 0.084 (mmol/mol SrCa)/°C and -0.23permil δ^{18}O /°C, which is consistent with interannual calibrations [5] (see also Section SI-3). The use of common calibrations that are confirmed by spatial calibrations and lab-experiments avoids issues associated with local and temporal calibrations [6]. Note, that the recalibration also affects the reported errors and replicate statistics.

SI-3. Reliability of coral-based estimates of marine variability

There are several classes of uncertainties associated with coral proxy records of sea surface temperature variability. For δ^{18}O proxies, the δ^{18}O of coral skeletons are temperature sensitive but are also influenced by the δ^{18}O of ambient seawater. Thus, changes in lateral and horizontal advection in the near surface ocean and changes in the balance of precipitation and evaporation will affect the signal. Our selection of δ^{18}O proxies reported to be mainly sensitive to temperature reduces but cannot wholly exclude this effect, perhaps especially on longer timescales [7]. The mechanistic relationship between Sr/Ca uptake and temperature involves numerous uncertainties. Sr/Ca uptake in corals may be affected by symbionts [8], growth rate [9], changes in the sea-water Sr/Ca, and the sampling process because of heterogeneity in coral ratios. See refs. [10] and [11] for more details.

Observational studies [5] indicate that seasonal signals in coral records can be attenuated because calcification occurs over a depth range that partially integrates across more than a single season. This attenuation appears to be limited to less-than-decadal timescales, however, and we use the calibration proporsed by Gagan et al. [5] as appropriate for decadal and longer timescales. In particular, for Sr/Ca we use 0.084 (mmol/mol SrCa)/°C and for δ^{18}O we use -0.23permil/°C. The implication is that we may somewhat underestimate interannual temperature variability using this calibration, which appears to be the case in comparisons between coral and instrumental SST spectral estimates (Fig. 3B, inset, Fig. S4). The approach of a single calibration for each coral proxy type is confirmed by comparing the spatial variance pattern of the coral records and instrumental SST. The recalibrated records correlate better to the instrumental SST variance pattern (R=0.84 recalibrated) than when using the individual calibrations published along with each record (R=0.26).

To further examine the calibration and the skill of the corals in recording SST variability, we compare the variability of coral SSTs against instrumental SSTs across different sites on interannual and interdecadal timescales (Fig. S4). There is a good correspondence between these independent estimates of temperature variance, showing a correlation of 0.93 on interannual and 0.53 on interdecadal timescales, but it is also useful to discuss those records for which there are disagreements in more detail. One outlier is the Sr/Ca record of Kilbourne et al. (2008) [12] that shows significantly more variability than the instruments. Another outlier is the proxy record reported by Dunbar et al. (1994) [13] that shows lower than instrumental variance, possibly reflecting the sparse instrumental measurements available in the earlier part of the century.

To test the sensitivity of our results to uncertainties in coral-based proxy estimates of marine temperature variability, we recalculate the spectra using only Sr/Ca coral records as well as only using the five corals most consistent with the instrumental variability at interdecadal timescales (Fig. S5). The SST variability reconstructed only from Sr/Ca (6 records) is very close to the SST variability using all records. The variability reconstructed from the five records that are closer to the decadal instrumental SST variance (Fig. S4) shows somewhat less variability but a similar scaling behavior and these

Reserved for Publication Footnotes

estimates remain much greater than the GCM estimates at the corresponding locations.

SI-4. Signal-to-noise estimates

One class of explanations for the model-data mismatch discussed in the main manuscript has to do with proxy noise, though this appears an unsatisfactory explanation. Such an explanation would require that proxy noise have an order of magnitude or more the variance in the proxy records as does the actual marine temperature variability signal. Below, two approaches are offered for estimating the ratio of temperature-to-noise variance, referred to as signal-to-noise ratios (SNR), and both show that the SNR is close to one.

SI-4.1 Forward model approach. The forward model of the signal and noise component described in [14] can be used to obtain an estimate of signal-to-noise ratios (SNR) for Mg/Ca and Uk37 based on our knowledge of the sampling process and instrumental temperature variability. The procedure that we use is to generate random timeseries following the spectral model with $\beta = 1$, corrupt them using the sampling regime found in the individual cores [14], and then apply the core-specific noise level, σ, and bioturbation with $\delta = 10$cm. Comparing the corrupted time-series and the original time-series after resampling both to a 250yr resolution results in a SNR estimate. An SNR of 0.5 is found as an average across all Mg/Ca records and 1.6 for Uk37. The lower proportion of signal relative to noise found for Mg/Ca records is in keeping with the greater variance correction that is applied to these records (Fig. 3).

SI-4.2 Correlation approach. An independent estimate of the SNR can be obtained from a comparison of nearby cores. Given a pair of cores with perfect time control and a signal shared entirely in common and each containing noise independent of the other, the SNR is simply $R/(1-R)$ where R is the correlation coefficient between both time series [15]. A complication, however, is that in our dataset most pairs of cores are separated in space such that the signal component cannot be expected to be closely correlated. To account for this effect we use climate model simulations to estimate the covariance of the signal component at each pair of core site locations as described further below. It is also possible that errors would be correlated, tending to bias the SNR high, but we have no evidence that this is the case. Further, temperature time series derived from sediment cores have a relative timing that is uncertain, tending to decrease the sample correlation coefficient and bias the SNR estimate low, but we do not account for either timing error or correlated noise in our estimates.

All pairs of cores that are less than 5000 km apart are examined. 5000 km approximates the centennial timescale decorrelation decay length, chosen as being somewhat larger than the 3800km decay length estimated at decadal timescales [16] under the expectation that longer timescales of variability also have larger spatial scales and because this affords sufficient samples for the statistical analysis. Note that we also include records with a mean temporal resolution of less than 150 years (Table S3) because restricting the analysis to cores with less than 100 year average sample resolution would give only three pairs of Mg/Ca separated by less that 5000 km. Every timeseries is resampled to a 250 year resolution and linearly detrended, thereby giving a consistent treatment with respect to the spectral analysis procedure.

The mean correlation of all pairs of each proxy type are computed. The null hypothesis of no relationship between records can be rejected for both Uk37 (p=0.01) and Mg/Ca records (p=0.02), which then strongly indicates that a common signal is present. To examine whether the forward modeling results presented earlier are consistent with the observed proxy correlation, we sample the 6000 year GCM simulation (orbital only) at the position of the cores using 250 year block averages. We add white noise to every model time series in order to obtain SNR between 0.1 and 10. Repeating this procedure 1000 times gives a distribution of mean correlations for the chosen SNR. An estimate of the SNR is then obtained by matching the correlation obtained from the model plus noise timeseries and the sample proxy correlation. This procedure results in a best estimate of 1.2 SNR for Mg/Ca and 1.4 SNR for Uk37. The difference between the Mg/Ca and Uk37 SNRs is qualitatively in the same direction as in the previous estimate from the forward model, though the Mg/Ca estimate is higher and the Uk37 estimate lower.

It is also possible to evaluate the uncertainty in the SNR estimates from correlation. We find that the proxy correlation is well inside the 95% range of the model plus noise correlation distribution when choosing the noise level to match an SNR of 0.5 for Mg/Ca and 1.6 for Uk37 (Fig. S1). This demonstrates consistency between our independent estimates of SNR derived variously from correlations and the forward spectral correction algorithm.

A number of choices are made in the foregoing estimates of SNR including those associated with bioturbation intervals, model derived estimates of covariance, and interpolation intervals. Although a full discussion of the robustness of these result could be provided, it suffices to say that values of SNR having a magnitude near one are consistently arrived at. SNR magnitudes of less than 0.1 would be needed to bring the proxy-observed and GCM-simulated marine variability into consistency with one another. Therefore, we conclude that both independent methodologies for estimating SNR indicate that the model-data mismatch in variance is not ascribable to noise.

SI-5. Sensitivity tests

The analysis presented here necessitates the selection of certain parameters and models. Here we discuss how the sensitivity of our results to these selections are small relative to the model-data discrepancy that is identified and that the consistency between the various proxy estimates of marine temperature variability is robust.

SI-5.1 Depth and seasonality. Foraminifera and alkenone producers do not evenly record the seasonal cycle nor do they monitor conditions exactly at the sea surface. For a more detailed discussion regarding seasonal and habitat influence on the recording of temperature see ref. [17] and references therein. To examine the influence of recording specific seasons or depths we additionally analysed GCM-produced variability in summer (JJA in the Northern Hemisphere, DJF in the Southern Hemisphere) and winter (DJF in NH, JJA in SH) as well as at 27m depth. Results show a small increase in variability at all timescales when considering seasonal temperature or considering subsurface temperature (Fig. S6). The latter is caused by the greater vertical advection of temperature anomalies. However, the scaling behavior of the variability is not affected and the magnitude of the increase is small relative to the model-proxy mismatch.

SI-5.2 Bioturbation depth assumption. Assuming no bioturbation leads to a small decrease in the estimated proxy SST variability of -18% for Uk37 and -19% for Mg/Ca. Assum-

ing that bioturbation acts over a 20cm vertical scale increases the estimated SST variability by 32% for Uk37 and 49% for Mg/Ca. Again, these changes are small relative to the order-of-magnitude discrepancies between models and data. For both Uk37 and Mg/Ca, the misfit between observed and estimated spectra in the spectral correction process [14] is smallest when assuming a 10cm bioturbation width.

SI-5.3 Inclusion of more records. Only those records having an average sampling interval of less than 100 years were included. As discussed in [14], it is difficult to correct for sampling and measurement noise in more coarsely resolved records. Nonetheless, accepting records that have a 150 year average sampling interval (Table S3) gives 4 more Mg/Ca and 5 more Uk37 records and does not change the main results (Fig. S8) but does lead to more scatter in variance ratios (i.e. those shown in Fig. 5), as expected from the Monte Carlo experiments [14].

1. Black DE, et al. (2007) An 8-century tropical Atlantic SST record from the Cariaco Basin: Baseline variability, twentieth-century warming, and Atlantic hurricane frequency. Paleoceanography 22.
2. Mueller PJ, Kirst G, Ruhland G, Von Storch I, Rosell-Mel A (1998) Calibration of the alkenone paleotemperature index U37K'based on core-tops from the eastern South Atlantic and the global ocean (60 N-60 S). Geochimica et Cosmochimica Acta 62:17571772.
3. Dekens P, Lea DW, Pak DK, Spero HJ (2001) Core top calibration of Mg/Ca in tropical foraminifera: refining paleo-temperature estimation (University of California, Santa Barbara).
4. Saenger C, Cohen AL, Oppo DW, Halley RB, Carilli JE (2009) Surface-temperature trends and variability in the low-latitude North Atlantic since 1552. Nature Geosci 2:492–495.
5. Gagan M, Dunbar G, Suzuki A (2012) The effect of skeletal mass accumulation in Porites on coral Sr/Ca and d18O paleothermometry. Paleoceanography 27:PA1203.
6. Osborn TJ (2004) CLIMATE: The Real Color of Climate Change? Science 306:621–622.
7. Thompson D, Ault T, Evans M, Cole J, Emile-Geay J (2011) Comparison of observed and simulated tropical climate trends using a forward model of coral d18O. Geophysical Research Letters 38:L14706.
8. Cohen AL (2002) The Effect of Algal Symbionts on the Accuracy of Sr/Ca Paleotemperatures from Coral. Science 296:331–333.
9. Goodkin N, Hughen K, Curry W, Doney S, Ostermann D (2008) Sea surface temperature and salinity variability at Bermuda during the end of the Little Ice Age. Paleoceanography 23:13 PP.
10. Correge T (2006) Sea surface temperature and salinity reconstruction from coral geochemical tracers. Palaeogeography, Palaeoclimatology, Palaeoecology 232:408–428.
11. Scott RB, Holland CL, Quinn TM (2010) Multidecadal Trends in Instrumental SST and Coral Proxy Sr/Ca Records. Journal of Climate 23:1017–1033.
12. Kilbourne KH, et al. (2008) Paleoclimate proxy perspective on Caribbean climate since the year 1751: Evidence of cooler temperatures and multidecadal variability. Paleoceanography 23:3220.
13. Dunbar RB, Wellington GM, Colgan MW, Glynn PW (1994) Eastern Pacific sea surface temperature since 1600 AD: The d18O record of climate variability in Galapagos corals. Paleoceanography 9:291315.
14. Laepple T, Huybers P (2013) Reconciling discrepancies between uk37 and mg/ca reconstructions of holocene marine temperature variability. Earth and Planetary Science Letters 375:418–429.
15. Fisher DA, Reeh N, Clausen HB (1985) Stratigraphic noise in the time series derived from ice cores. Annals of Glaciology 7:7683.
16. Jones PD, Osborn TJ, Briffa KR (1997) Estimating sampling errors in large-scale temperature averages. Journal of Climate 10:25482568.
17. Lohmann G, Pfeiffer M, Laepple T, Leduc G, Kim JH (2012) A model-data comparison of the Holocene global sea surface temperature evolution. Climate of the Past Discussions 8:10051056.
18. Quinn TM, et al. (1998) A multicentury stable isotope record from a New Caledonia coral: Interannual and decadal sea surface temperature variability in the southwest Pacific since 1657 A.D. Paleoceanography 13:PP. 412–426.
19. Asami R, et al. (2005) Interannual and decadal variability of the western Pacific sea surface condition for the years 1787-2000: Reconstruction based on stable isotope record from a Guam coral. Journal of Geophysical Research.
20. Hendy EJ (2002) Abrupt Decrease in Tropical Pacific Sea Surface Salinity at End of Little Ice Age. Science 295:1511–1514.
21. Linsley BK (2000) Decadal Sea Surface Temperature Variability in the Subtropical South Pacific from 1726 to 1997 A.D. Science 290:1145–1148.
22. Calvo E, Pelejero C, De Deckker P, Logan G (2007) Antarctic deglacial pattern in a 30 kyr record of sea surface temperature offshore South Australia. Geophysical research letters 34:L13707.
23. Linsley BK, et al. (2004) Geochemical evidence from corals for changes in the amplitude and spatial pattern of south pacific interdecadal climate variability over the last 300 years. Climate Dynamics 22:1–11.
24. Weldeab S, Lea DW, Schneider RR, Andersen N (2007) 155,000 Years of West African Monsoon and Ocean Thermal Evolution. Science 316:1303–1307.
25. Stott L, et al. (2004) Decline of surface temperature and salinity in the western tropical Pacific Ocean in the Holocene epoch. Nature 431:56–59.
26. Farmer EJ, Chapman MR, Andrews JE (2008) Centennial-scale Holocene North Atlantic surface temperatures from Mg/Ca ratios in Globigerina bulloides. Geochemistry Geophysics Geosystems 9.
27. Cleroux C, et al. (2012) High-resolution sea surface reconstructions off Cape Hatteras over the last 10ka. Paleoceanography 27:14 PP.
28. Marchitto TM, Muscheler R, Ortiz JD, Carriquiry JD, van Geen A (2010) Dynamical Response of the Tropical Pacific Ocean to Solar Forcing During the Early Holocene. Science 330:1378–1381.
29. Doose-Rolinski H, Rogalla U, Scheeder G, Lueckge A, Rad Uv (2001) High-resolution temperature and evaporation changes during the Late Holocene in the northeastern Arabian Sea. Paleoceanography 16:P. 358.
30. Kim J, et al. (2007) Impacts of the North Atlantic gyre circulation on Holocene climate off northwest Africa. Geology 35:387 –390.
31. Zhao M, Huang C, Wang C, Wei G (2006) A millennial-scale U37K sea-surface temperature record from the South China Sea (8N) over the last 150 kyr: Monsoon and sea-level influence. Palaeogeography, Palaeoclimatology, Palaeoecology 236:39–55.
32. Kim J, et al. (2004) North Pacific and North Atlantic sea-surface temperature variability during the Holocene. Quaternary Science Reviews 23:2141–2154.
33. Emeis K, Struck U, Blanz T, Kohly A, Vo M (2003) Salinity changes in the central Baltic Sea (NW Europe) over the last 10000 years. The Holocene 13:411 –421.
34. Sachs J (2007) Cooling of Northwest Atlantic slope waters during the Holocene. Geophysical research letters 34:L03609.
35. Isono D, et al. (2009) The 1500-year climate oscillation in the midlatitude North Pacific during the Holocene. Geology 37:591 –594.
36. Calvo E, Grimalt J, Jansen E (2002) High resolution U37K sea surface temperature reconstruction in the Norwegian Sea during the Holocene. Quaternary Science Reviews 21:1385–1394.
37. Harada N, et al. (2006) Rapid fluctuation of alkenone temperature in the southwestern Okhotsk Sea during the past 120 ky. Global and Planetary Change 53:29–46.
38. Rodrigues T, Grimalt JO, Abrantes FG, Flores JA, Lebreiro SM (2009) Holocene interdependences of changes in sea surface temperature, productivity, and fluvial inputs in the Iberian continental shelf (Tagus mud patch). Geochemistry Geophysics Geosystems 10.
39. Bendle J, Rosell-Mele A (2007) High-resolution alkenone sea surface temperature variability on the North Icelandic Shelf: implications for Nordic Seas palaeoclimatic development during the Holocene. The Holocene 17:9.
40. Lamy F, Rhlemann C, Hebbeln D, Wefer G (2002) High- and low-latitude climate control on the position of the southern Peru-Chile Current during the Holocene. Paleoceanography 17:10 PP.
41. Lueckge A, et al. (2009) Monsoon versus ocean circulation controls on paleoenvironmental conditions off southern Sumatra during the past 300,000 years. Paleoceanography 24:14 PP.
42. Jungclaus JH, et al. (2010) Climate and carbon-cycle variability over the last millennium. Clim. Past 6:723–737.
43. Fischer N, Jungclaus J (2011) Evolution of the seasonal temperature cycle in a transient Holocene simulation: orbital forcing and sea-ice. Climate of the Past 7:1139–1148.
44. Schmidt GA, et al. (2011) Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). Geoscientific Model Development 4:33–45.
45. Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93:485–498.
46. Braconnot P, et al. (2012) Evaluation of climate models using palaeoclimatic data. Nature Climate Change 2:417–424.
47. Liu Z, et al. (2009) Transient simulation of last deglaciation with a new mechanism for bolling-allerod warming. Science 325:310–314.
48. Farmer E (2005) Holocene and deglacial ocean temperature variability in the Benguela upwelling region: implications for lowlatitude atmospheric circulation. Paleoceanography 20.
49. Sun Y, Oppo D, Xiang R, Liu W, Gao S (2005) Last deglaciation in the Okinawa Trough: Subtropical northwest Pacific link to Northern Hemisphere and tropical climate. Paleoceanography 20:9 PP.
50. Xu J, Holbourn A, Kuhnt W, Jian Z, Kawamura H (2008) Changes in the thermocline structure of the Indonesian outflow during Terminations I and II. Earth and Planetary Science Letters 273:152–162.
51. Thornalley DJR, Elderfield H, McCave IN (2009) Holocene oscillations in temperature and salinity of the surface subpolar North Atlantic. Nature 457:711–714.
52. Marchal O, et al. (2002) Apparent long-term cooling of the sea surface in the northeast Atlantic and Mediterranean during the Holocene. Quaternary Science Reviews 21:455–483.

53. Zhao M, Beveridge NAS, Shackleton NJ, Sarnthein M, Eglinton G (1995) Molecular stratigraphy of cores off northwest Africa: Sea surface temperature history over the last 80 Ka. Paleoceanography 10:P. 661.

54. deMenocal P, Ortiz J, Guilderson T, Sarnthein M (2000) Coherent High- and Low-Latitude Climate Variability During the Holocene Warm Period. Science 288:2198–2202.

55. Barron JA, Heusser L, Herbert T, Lyle M (2003) High-resolution climatic evolution of coastal northern California during the past 16,000 years. Paleoceanography 18:19 PP.

56. Cacho I, et al. (2001) Variability of the western Mediterranean Sea surface temperature during the last 25,000 years and its connection with the Northern Hemisphere climatic changes. Paleoceanography 16:PP. 40–52.

57. Schmidt GA, et al. (2014) Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. Journal of Advances in Modeling Earth Systems 6:141–184.

58. Gao C, Robock A, Ammann C (2008) Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. Journal of Geophysical Research 113:15 PP.

59. Crowley TJ, et al. (2008) Volcanism and the little ice age. PAGES news 16:2223.

Table S1. Proxy data used in the main study

Name	Ref.	Lat. $^{\circ}N$	Lon. $^{\circ}E$	sed. rate (cm/kyr)	mean Δt (yr)	interp. Δt (yr)	duration (yr)	proxy	error 1sd ($^{\circ}C$)
Caledonia	[18]	-22.5	166.5		1	1	333	δ^{18}O	0.22
Galapagos	[13]	-0.4	-91.2		1	1	346	δ^{18}O	0.29
Guam	[19]	13.6	144.8		1	1	208	δ^{18}O	0.10
Gr.Barrier Reef	[20]	-18.0	146.5		5	5	415	Sr/Ca	0.06
Rarotonga 2R	[21]	-21.5	-159.5		1	1	268	Sr/Ca	0.17
Bermuda	[9]	32.0	296.0		1	1	216	Sr/Ca	0.11
Flinders reef	[22]	-17.5	148.3		5	5	280	Sr/Ca	0.11
Turrumote	[12]	17.9	293.0		1	1	253	Sr/Ca	0.30
Fiji 1F	[23]	-16.8	179.2		1	1	215	Sr/Ca	0.17
Bahamas	[4]	25.8	-78.6		1	5	439	growth rate	0.07
Cariaco Mg/Ca	[1]	10.8	295.2		1.4	2	769	Mg/Ca G.bulloides	0.17
MD03-2707	[24]	2.5	9.4	55	36	100	6600	Mg/Ca G.ruber (pink)	
MD98-2176	[25]	-5.0	133.4	50	43	100	6818	Mg/Ca G.ruber (white)	
MD99-2155	[26]	57.4	-27.9	166	52	100	6440	Mg/Ca G.bulloides	
MD99-2203	[27]	35.0	-75.2	53	19	100	6374	Mg/Ca G.ruber (white)	
MD98-2181	[25]	6.3	125.8	80	43	150	6969	Mg/Ca G.ruber (white)	
MV99-GC41/PC14	[28]	25.2	247.3	79	90	150	6091	Mg/Ca G.bulloides	
SO90-39KG/56KA	[29]	24.8	65.9	123	20	100	4880	Uk'37	
GeoB6007	[30]	30.9	-10.3	65	31	100	6750	Uk'37	
MD97-2151	[31]	8.7	109.9	39	49	100	6020	Uk'37	
SSDP-102	[32]	35.0	128.9	216	61	150	6870	Uk'37	
IOW 225514	[33]	57.8	8.7	66	72	150	5980	Uk'37	
CH07-98-GGC19	[34]	36.9	-74.6	27	68	150	6470	Uk'37	
OCE326-GGC30	[34]	43.9	-62.8	30	72	150	6940	Uk'37	
KR02-06	[35]	36.0	141.8	30	51	150	7013	Uk'37	
MD952011	[36]	67.0	7.6	74	58	200	6450	UK37	
MD01-2412	[37]	44.5	145.0	95	73	200	6920	Uk'37	
D13882	[38]	38.6	-9.5	55	53	200	6530	Uk'37	
IOW 225517	[33]	57.7	7.1	52	94	200	5170	Uk'37	
JR51-GC35	[39]	67.0	-18.0	48	98	200	6880	UK37	
GeoB 3313-1	[40]	-41.0	-74.5	107	90	200	6930	Uk'37	
GeoB 5901-2	[30]	36.4	-7.1	13	80	200	5840	Uk'37	
SO139-74KL	[41]	-6.5	103.8	106	78	200	5870	Uk'37	

Errors reported above are relative to the interpolated resolution and are converted to ($^{\circ}C$) using the calibrations described in (Sec. SI-2). For the Guam coral, only the variability from an internal standard was reported and the mean replicate variability from the two other δ^{18}O coral studies (0.0835 permil) is assumed. For Flinders Reef a 0.1% error is assumed because the reported 0.02% only represent the analytical error. For Fiji 1F and Rarotonga 2R we understood that uncertainty rates were of 0.48% (personnel communication, Braddock Linsley 05 Jun 2012) and later that they were 0.15%, in accord with values published in Linsley et al. (2000) and Linsley et al. (2004) (personnel communication, Braddock Linsley 18 Oct 2014). We have assumed the larger uncertainty rate in all calculations, though have also calculated results for the smaller 0.15% rate, which would lead to less than a 1% change in the overall coral variance estimate, being thus unimportant for our particular results. Sedimentation rates are mean values over the last 7kyr or over the duration of the record, whichever is shorter. Duration is the length of the record inside the 7kyr BP to present day interval used in this study where BP is with respect to 1950 AD.

Table S2. Model simulations

Name	No.	Model components	Forcing	Resolution	Time	Ref.
Millennium-Forced	5	ECHAM5 (atmosphere) MPIOM (ocean) HAMOCC5 (biogeochem.) JSBACH (land)	solar, volcanic aerosols, GHG (CO2 interactive), land-cover, orbital	atmosphere (T31), ocean (22-350km)	800-2005 CE	[42]
Millennium-CTRL	1	ECHAM5 (atmosphere) MPIOM (ocean) HAMOCC5 (biogeochem.) JSBACH (land)	preindustial boundary conditions, no forcing	atmosphere (T31), ocean (22-350km)	3100yr	[42]
Orbital	1	ECHAM5 (atmosphere) MPIOM (ocean) JSBACH (land) +interactive vegetation	orbital	atmosphere (T31), ocean (22-350km)	4000BCE - 2000 CE	[43]
CMIP5/PMIP3 past1000	15	coupled ocean atmosphere models	full forcing [44]	varies	850-1850CE	[45, 46]
TraCE-21ka	1	CCSM3	GHG, orbital, ice sheets,paleogeography, meltwater	T31gx3v5	last 21kyr	[47]

Table S3. Proxy data $\Delta t \geq$100yr, used in Sec. SI4.2 and SI5.3

Name	Ref.	Lat. $^\circ N$	Lon. $^\circ E$	sed. rate (cm/kyr)	mean Δt (yr)	interp. Δt (yr)	duration (yr)	proxy
ODP1084B	[48]	-25.5	13.3	9	126	200	6916	Mg/Ca G. bulloides
A7	[49]	27.8	127.0	11	125	200	5866	Mg/Ca G. ruber
MD01-2378	[50]	-13.1	121.8	22	131	200	6300	Mg/Ca G. ruber
RAPID-12-1K	[51]	62.1	-17.8	21	100	200	6909	Mg/Ca G. bulloides
MD952015	[52]	58.8	-26.0	50	101	200	6260	UK'37
ODP 658C	[53, 54]	20.8	-18.6	21	100	200	6900	UK'37
ODP 1019C	[55]	41.7	-124.9	40	132	200	6840	UK'37
MD95-2043	[56]	36.1	-2.6	37	133	200	5980	UK'37
OCE326-GGC26	[34]	43.5	-54.9	30	128	200	6770	UK'37

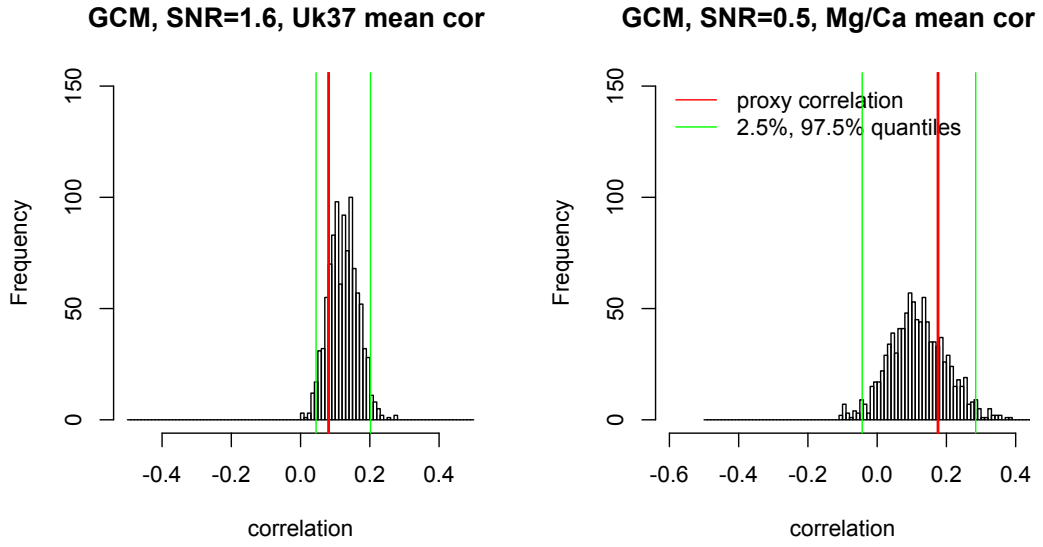

Fig. S1. Check of consistency between the forward model and correlation based approaches for estimating signal-to-noise ratios (SNRs). To obtain SNRs equivalent to those estimated from the forward model for Mg/Ca and Uk37, white noise is added to GCM temperature simulations at the proxy positions. The observed average proxy correlations, indicated by a vertical red line, are inside the distribution of the simulated correlations, showing consistency between the two different estimates of SNR. The 2.5 and 97.5 quantiles of the simulated correlations are indicated by green vertical lines.

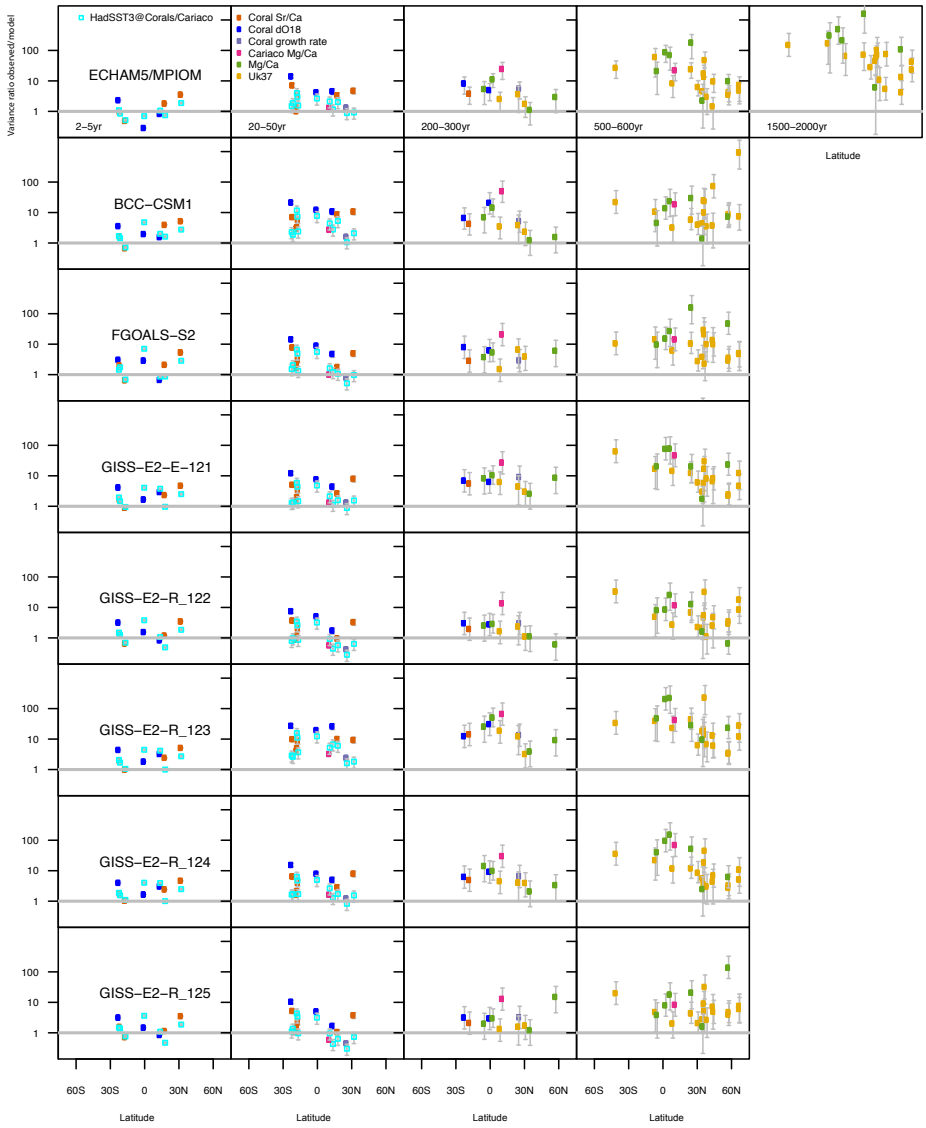

Fig. S2. Latitudinal dependence of the model-data mismatch at different timescales for the CMIP5/PMIP3 past1000 simulations. See Fig. 5 for the full caption. In the first row, ECHAM5/MPIOM is shown as a reference. As the past1000 simulations contain only the last millenium, the millenial variance ratio could not be estimated.

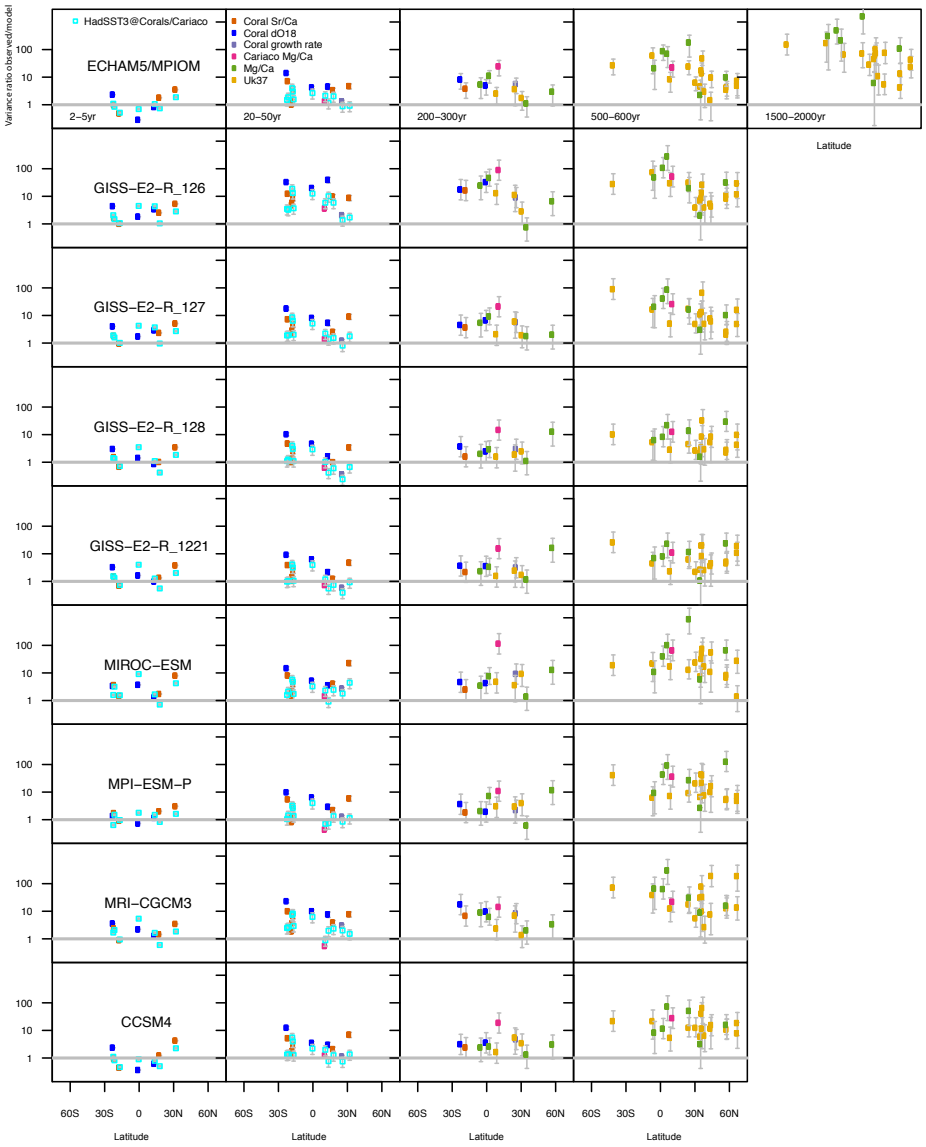

Fig. S3. Latitudinal dependence of the model-data mismatch at different timescales for the CMIP5/PMIP3 past1000 simulations, continuation

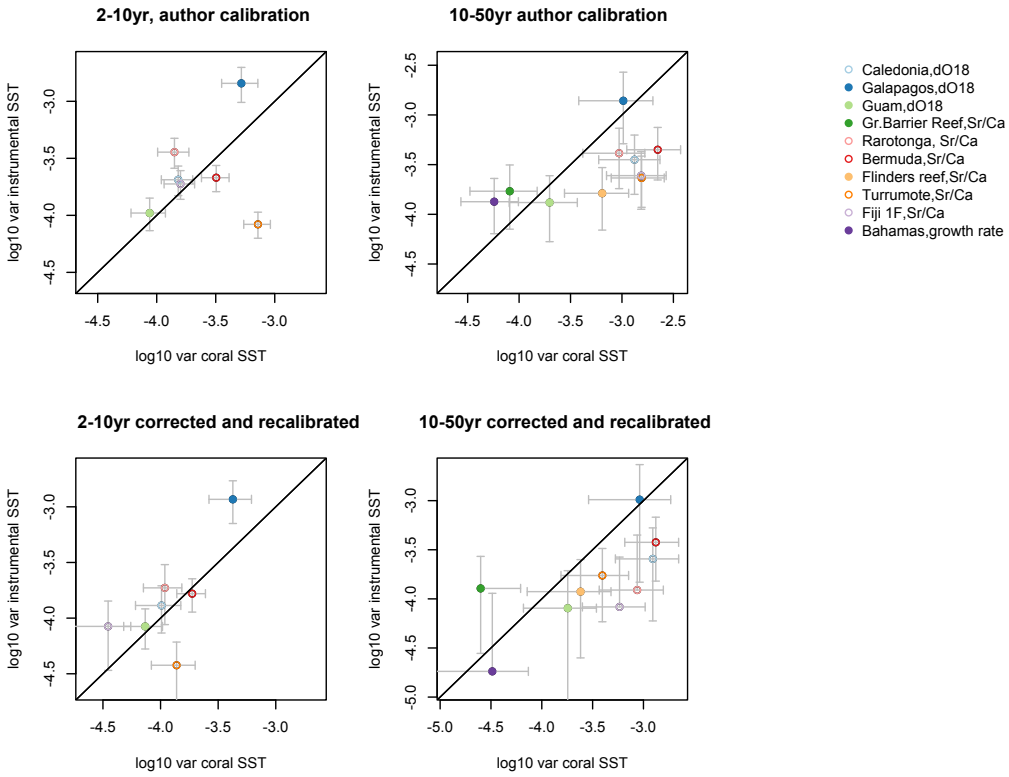

Fig. S4. Comparison of instrumental (HadSST3) and coral based SST variance for interannual (left column) and interdecadal (right column) timescales. Top row: uncorrected instrumental and proxy data using the authors calibration; lower row: variances corrected for sampling and measurement error using the common calibration. Using the common calibration increases the correlation between instrumental and coral SST variance. On interannual timescales coral proxies tend to underestimate variability, as expected from the selected calibration [5]. At interdecadal timescales some corals tend to overestimate variability, possibly because the biological-smoothing effect was underestimated or because proxy noise has been underestimated or measurement noise was overestimated. The records most consistent with the instrumental data regarding interdecadal variability are marked as filled dots and analysed separately in a sensitivity experiment (Fig. S5).

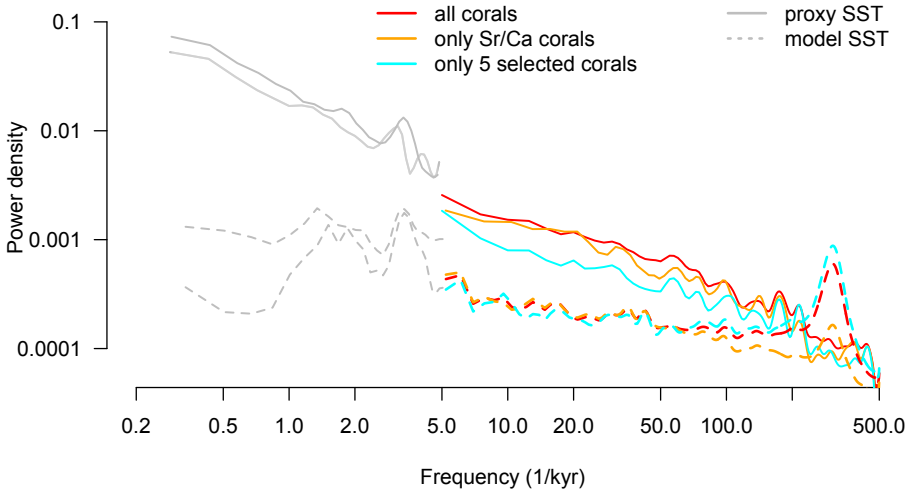

Fig. S5. Sensitivity of spectral results to the choice of coral proxies. As main Fig. 3B but only using Sr/Ca records or only using the five records that are closest in variability to their instrumental counterparts (Fig. S4). In addition, the GCM SST spectrum, sampled at the coral position is shown. In either case, the corals show a scaling similar to that of the entire collection and much greater variability than the GCM results.

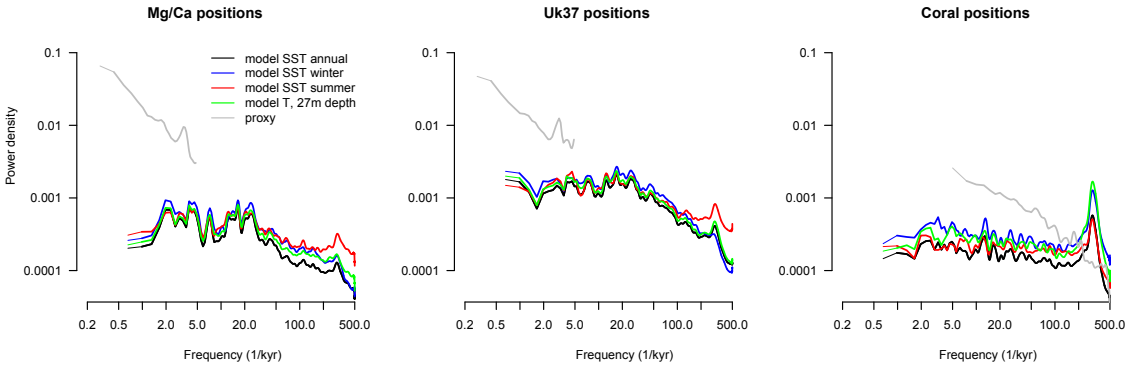

Fig. S6. Dependence of marine temperature variability on season and water depth in the ECHAM5/MPIOM unforced control simulation. Shown is variability for annual, summer, and winter surface conditions as well as annual conditions at 27m depth at the Mg/Ca core site positions (left panel) and the Uk37 core sites (right panel). A small increase in variability is observed in each case relative to the annual average surface conditions that is largely frequency independent. Depth has a greater influence at the Mg/Ca than Uk37 core positions because of the shallower mixed layer depth in the tropics where many of the Mg/Ca records are located.

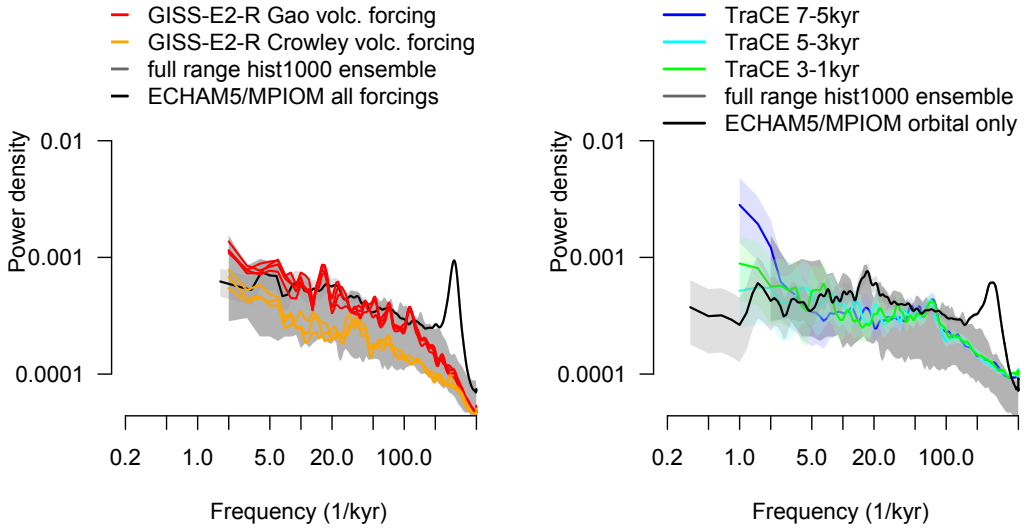

Fig. S7. Dependence of simulated marine temperature variability on the forcing. Volcanic forcing (left panel): GISS-E2-R [57] ensemble members p122,p125,p128,p1221 using the stronger Gao et al., volcanic forcing [58] are on the upper end of the CMIP5/PMIP3 model envelope (grey) whereas the ensemble members from the same model but using the Crowley et al, [59] forcing are on the lower side of the envelope. Freshwater forcing (right panel): The 7-5kyr BP section of the TraCE-21ka model simulation [47], which includes freshwater forcing, shows more variability on multicentennial and millenial timescales than later time periods without freshwater forcing

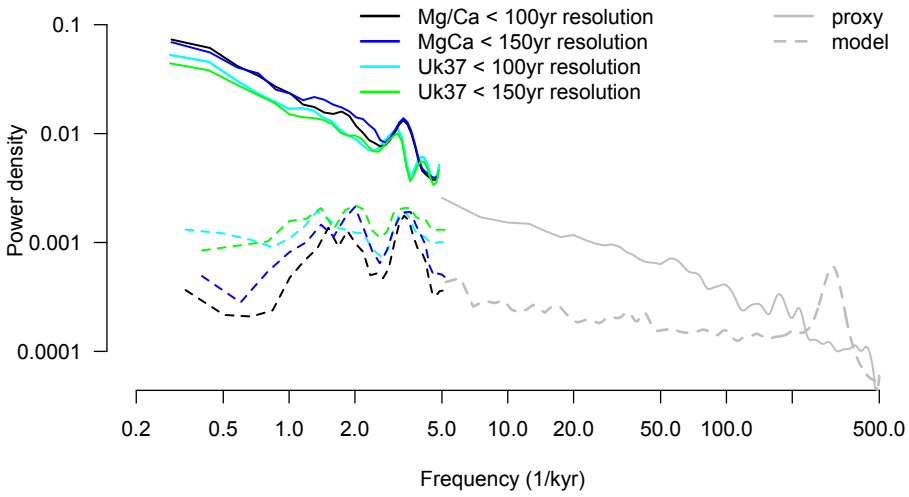

Fig. S8. Sensitivity of average spectra to inclusion of lower-resolution proxy records. Proxy spectra are after filtering for measurement and sampling errors and GCM spectra are at the proxy positions for records with a mean resolution less than 100 years and including records with an average resolution of up to 150 years (Table S3). There is little difference when these additional records are included.