nature geoscience

Perspective

https://doi.org/10.1038/s41561-023-01299-9

# Regional but not global temperature variability underestimated by climate models at supradecadal timescales

Received: 19 January 2023

Accepted: 21 September 2023

Published online: 06 November 2023

Check for updates

T. Laepple[1,2] ✉, E. Ziegler[3,4], N. Weitzel[3], R. Hébert[1], B. Ellerhoff[3,4,13], P. Schoch[5], B. Martrat[6], O. Bothe[7], E. Moreno-Chamarro[8], M. Chevalier[9,10], A. Herbert[11,12] & K. Rehfeld[3,4]

Knowledge of the characteristics of natural climate variability is vital when assessing the range of plausible future climate trajectories in the next decades to centuries. The reliable detection of climate fluctuations on multidecadal to centennial timescales depends on proxy reconstructions and model simulations, as the instrumental record extends back only a few decades in most parts of the world. Systematic comparisons between model-simulated and proxy-based inferences of natural variability, however, often seem contradictory. Locally, simulated temperature variability is consistently smaller on multidecadal and longer timescales than is indicated by proxy-based reconstructions, implying that climate models or proxy interpretations might have deficiencies. In contrast, at global scales, studies found agreement between simulated and proxy reconstructed temperature variations. Here we review the evidence regarding the scale of natural temperature variability during recent millennia. We identify systematic reconstruction deficiencies that may contribute to differing local and global model–proxy agreement but conclude that they are probably insufficient to resolve such discrepancies. Instead, we argue that regional climate variations persisted for longer timescales than climate models simulating past climate states are able to reproduce. This would imply an underestimation of the regional variability on multidecadal and longer timescales and would bias climate projections and attribution studies. Thus, efforts are needed to improve the simulation of natural variability in climate models accompanied by further refining proxy-based inferences of variability.

Climate variability results from a forced component driven by the planetary energy imbalance and from chaotic variations linked to the internal dynamics of the climate system. While the current long-term increase in global mean temperature is mostly due to climate forcings related to human activities[1], internal variability (Box 1) dominates the regional, short-term changes that are most relevant to societies[2,3]. Knowledge of the climate system's variability is required to anticipate the full range of possible future climate change and to increase the impact of societal mitigation and adaptation efforts. While forced variability is relatively well understood[1], regional internal variability is largely uncertain, and society-relevant changes persisting for decades could be outside the range of current projections[4]. Climate models can provide direct estimates of internal variability at all scales[5]. However, uncertainties are introduced by processes not being included and models mostly being

A full list of affiliations appears at the end of the paper. ✉e-mail: tlaepple@awi.de

https://doi.org/10.1038/s41561-023-01299-9

## BOX 1

# Glossary of terms

**Spatial scales**: In this study, local refers to a single proxy record or model grid box and regional to any aggregation of multiple proxy records or grid boxes smaller than a hemisphere.

**Temporal scales**: Subdecadal denotes interannual to decadal timescales, and supradecadal multidecadal to millennial timescales.

**Spectrum**: For a given time series, the PSD (here referred to simply as spectrum) provides a measure of variability as a function of timescale. For temperature, this spectrum often follows a power law with a scaling coefficient that describes its temporal persistence. A larger scaling implies stronger correlations across timescales and longer persistence times.

**Temporal persistence**: The strength of (anti-)correlation of successive values in a time series.

**Internal climate variability**: Climate variability has a forced component and an unforced (that is, internal) component. Internal variability arises from nonlinear interactions between the various feedbacks and components in the climate system, from stochastic fluctuations and from deterministic chaos. As such, it appears in unforced 'control' model simulations. Forced variability results from external drivers of the climate system, irrespective of whether they are anthropogenic or natural. At first order, internal and forced variability are independent, but forced variability can influence the timing and strength of internal climate variability. Natural variability refers to internal and naturally forced.

**Climate model**: Unless otherwise stated, we use model as shorthand to denote global atmosphere/ocean general circulation models.

**Model–data comparison**: In the context of model–data comparison, we use data as shorthand to refer to both instrumental observations and proxy-based reconstructions of surface temperature, but not for results produced by numerical models.

validated against recent observational constraints[6]. Furthermore, computational constraints necessitate parameterizations and only allow small ensemble simulations on long timescales that are needed to constrain longer-than-decadal timescales (here termed supradecadal timescales). While instrumental data can be used to estimate the contribution of internal variability for periods of up to several decades, the limited length of these records and the concurrent changes in anthropogenic forcing imply that fluctuations on supradecadal timescales are poorly constrained. Thus, model simulations remain largely unvalidated for variability on multidecadal timescales and beyond. Approaches that exploit observational records are promising[7], but their robustness and scope need to be further improved using palaeoclimate reconstructions to achieve confidence on those timescales.

Current comparisons of temperature variability in model simulations and reconstructions or instrumental observations (hereafter model–data comparisons) seem contradictory. While some studies have found agreement between the variability simulated by climate models and the variability inferred from instrumental observations[8] or proxy-based reconstructions[9–11], others suggest disagreement for the instrumental period[12,13] and past climates[4,14,15].

Typical examples of global and local temperature reconstructions from proxy data and model simulations reveal that these contradictory findings are a result of the distinct nature of global and local climate variability (Fig. 1). Global mean temperature time series from proxy-based reconstructions (Past Global Changes (PAGES) 2k proxy temperature database[16]) and models (the Coupled Model Intercomparison Project Phase 5 (CMIP5)/Paleoclimate Modelling Intercomparison Project Phase III (PMIP3) and Coupled Model Intercomparison Project Phase 6 (CMIP6) last millennium ensemble[17], referred to here as CMIP5/6; Extended Data Table 1) show similar timings and amplitudes of the anomalies throughout most of the past millennium (Fig. 1a) and the power spectra of model simulations and reconstructions are consistent across most timescales (Fig. 1c). In contrast, local temperature variabilities are largely uncorrelated, as illustrated by the Cariaco sea surface temperature (SST) reconstruction[18] and simulated temperatures for the same location (Fig. 1b). The reconstruction exhibits larger amplitudes than individual simulations and even exceeds the ensemble's spread. On decadal and shorter timescales the proxy and model-based power spectra are more consistent (Fig. 1d), but they diverge for supradecadal timescales. There, model spectra flatten, while the local SST record shows a steep increase in the spectral density towards longer timescales. This behaviour is not limited to this example but can also be found in high-resolution records of the PAGES 2k compilation (ref. 19 and Extended Data Fig. 1), in lower-resolution SST records[14] and in land temperature records[4].

To systematically disentangle the structure of agreement and disagreement between model simulations and reconstructions, we performed a meta-study of the existing literature on surface temperature variability in the Holocene and past millennium, following the respective author's original assessments of model–data consistency (Fig. 2 and Extended Data Fig. 2; Methods). At the global scale, there is a high level of agreement on all timescales[8,10–12,16,19,20]. As the spatial scale decreases from global to hemispheric, the results are more mixed. For the Northern Hemisphere some studies found agreement on decadal to centennial timescales[9,11,21], while others reported weaker simulated variability than that reconstructed on interannual to multicentennial scales[22,23]. Owing to the sparsity of proxy coverage of the Southern Hemisphere, there is a clear bias towards the Northern Hemisphere in global and hemispheric studies. At regional and local scales (Box 1), the findings are strongly timescale-dependent. On interannual scales, most studies found model–data agreement[4,8,12,14,19,21–26] or reported mixed results[13,26–29]. Disagreement emerged on decadal[19] or multidecadal[8,12,14] scales. On supradecadal timescales, the majority of studies showed disagreement between reconstructions and models[4,12–14,19,21,25,27] with reconstructions yielding higher magnitudes of temperature variability than simulations.

This leaves us with the enigma of a model–data match at global scales and a mismatch at local and regional scales for supradecadal timescales. Possible causes for this could be shortcomings in the ability of current climate models to simulate local and regional temperature variability, or systematic deficiencies in global or local temperature reconstructions.

## Reconstruction deficiencies alone do not explain the scale enigma

Pre-instrumental temperature estimates are primarily reconstructed from physical and biogeochemical proxies measured in natural archives such as glacier ice, sediments and trees[30]. This proxy information is transformed into estimates of past temperatures using statistical methods. While many sources of uncertainties and biases have been identified in these statistical reconstructions[31–34], we focus here on potential underlying causes of the diagnosed scale enigma.

Studies that have found consistent global variability in reconstructions and simulations typically employed multi-proxy databases comprising predominantly annually resolved records. These have

https://doi.org/10.1038/s41561-023-01299-9

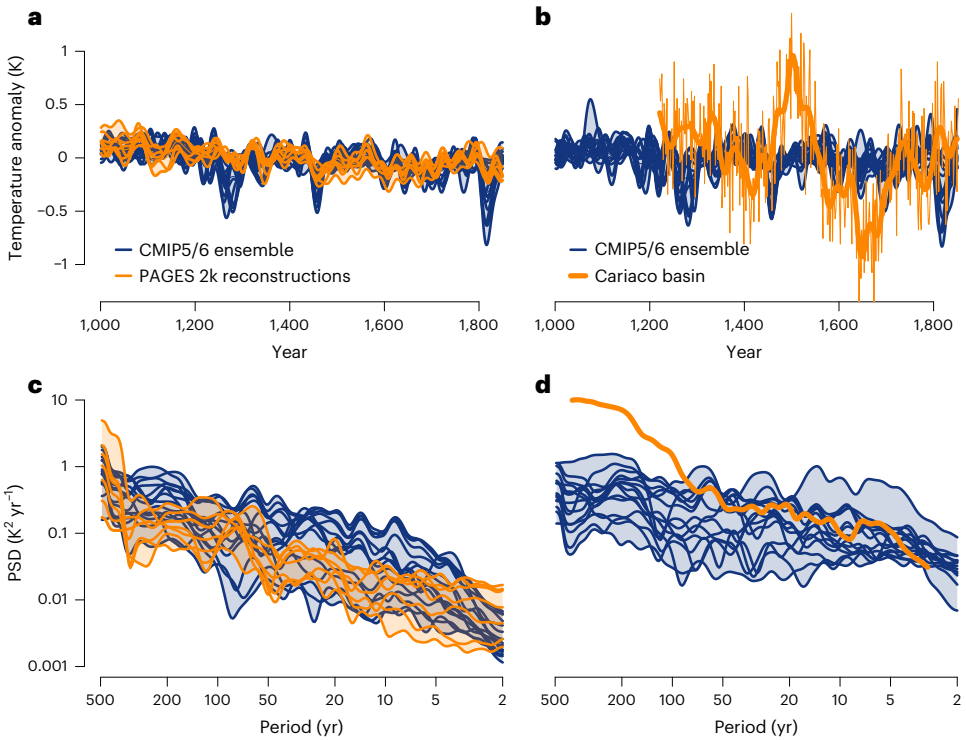



**Fig. 1 | Examples of local and global surface temperature variability in models and reconstructions over the 1,000–1,800 CE time span. a**, Time series of global mean surface temperature from CMIP5/6 model simulations and all PAGES 2k proxy reconstructions[16]. **b**, Time series of a local SST reconstruction (Cariaco[18]) and of model simulations for the same location (*y*-axis as in (**a**)). The thin lines show annual values, thick lines are low-pass filtered (cutoff frequency 1/30 yr) to focus on the supradecadal variations. **c**,**d**, Power spectral density (PSD) of the global (**c**) and local (**d**) simulated and reconstructed temperature time series on shared *y*-axis. Coloured shading indicates the range of the model and proxy reconstruction ensemble.

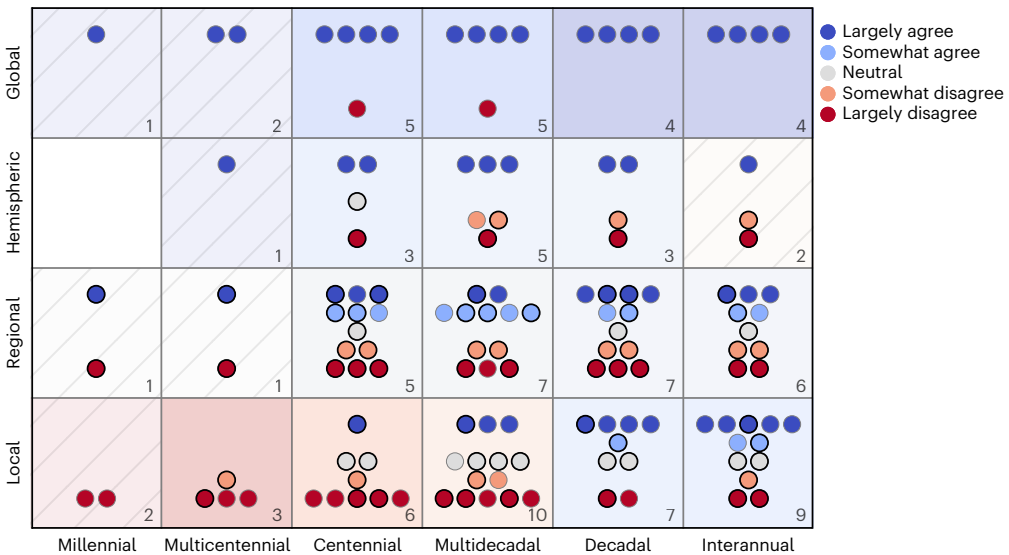



**Fig. 2 | Model–data (dis)agreement on Holocene temperature variability in the literature.** Agreement as a function of temporal (*x* axis) and spatial (*y* axis) scales. Each circle represents a specific study, colour coded according to the strength of (dis)agreement (Methods). Some studies report different rates of agreement at one spatiotemporal scale; that is, depending on properties such as region or proxy type. Such multiple occurrences in one box are marked by a thick line around the circle. The number of distinct studies is given at the bottom right of each box. Diagonal lines indicate a small number of studies (one or two) for the given spatiotemporal scale. See Extended Data Fig. 2 indicating the specific literature studies (refs. 4,8–14,16,19–29,97).

been calibrated using the period that overlaps with the instrumental record (temporal calibration). A number of studies have suggested that global temperature reconstructions might underestimate variability (such as ref. 34 and references therein). The temporal calibration step can underestimate supradecadal variability, particularly when the proxy is used as predictor (direct regression)[35]. Furthermore, current multi-proxy databases are dominated by tree ring records (for example ~60% in the PAGES 2k 2017 database) that can underestimate

supradecadal variability due to the removal of individual growth trends[36]. These arguments have sparked the development of new methods including indirect regression (which avoids the underestimation of low-frequency variability in the calibration step), ensemble methods and Bayesian models[37]. Pseudo-proxy experiments suggest that these new methods do not substantially underestimate supradecadal variability[31]. While the few reconstructions that do not rely on tree ring records exhibit the highest supradecadal variability[38,39] and differing methodologies can produce a spread of up to one order of magnitude, even the global reconstructions with the highest magnitude variability are still consistent with the ensemble of CMIP5/6 simulations (Fig. 1a,c).

Unlike global reconstructions, studies focusing on local scales often rely on proxies from sedimentary archives (lake and marine sediments, glacier ice) that are calibrated using independent data from laboratory experiments[40], physical models[41] or modern spatial relationships[42]. These calibrations are not subject to the underestimation of supradecadal variability found for direct temporal regression[35]. Instead, 'proxy noise' (that is, the variability enhancement from non-climatic processes) has been proposed as the most likely culprit for overestimations of supradecadal variability in these local reconstructions. Examples of such processes are aliasing of the seasonal cycle[43], spatial heterogeneity in the archive[44], recording of site-specific features such as upwelling events that are not representative of mean conditions[45], vital effects or measurement errors. As local variability estimates typically rely on single records, there can be no dampening of proxy noise through averaging. In contrast to these variability-enhancing mechanisms, other non-climatic processes such as bioturbation of sediments[46], isotopic diffusion in ice[47], slow response times of biological systems[48] and homeostasis[49] can lead to timescale-dependent underestimation of local variability. Biases from proxy noise and timescale-dependent temperature sensitivities can be minimized through statistical corrections[50] and the use of proxy system models[51]. These corrections lead to consistent estimates of variability across independent proxy types and with instrumental observations on overlapping timescales[4,14], strengthening the credibility of reconstructed variability. Nonetheless, corrected estimates still show the model–data variability mismatch on supradecadal scales[14,27]. Supporting the evidence in the literature, we found that local variability reconstructed from instrumentally calibrated annually resolved records is consistent with marine and terrestrial proxies and displays a similar model–data variability mismatch on supradecadal timescales (Extended Data Fig. 1).

Chronological uncertainties can complicate the comparison of individual events in proxies and simulations. However, age model errors have little influence on local variability estimates from single records[52]. While averaging across records with age uncertainty leads to an underestimation of variability[4], the age uncertainties for annually resolved records of the Common Era is usually below 10 years (ref. 53). Thus, we expect little to no effect on supradecadal global temperature variability estimates. It is often not possible to unambiguously attribute proxy variations to a specific physical variable (for example, trees can be sensitive to temperature or moisture) or season. However, the mismatch for local supradecadal variability would be even higher if we assumed that proxies represented hydroclimate, instead of temperature variations (Extended Data Fig. 3). In addition, the local model–data disagreement is also found in geochemical proxies that are only sensitive to temperature (for example, Mg/Ca ratios in foraminifera or Sr/Ca ratios in corals). Taken together, it is unlikely that chronological uncertainties or hydroclimate influences on proxies contribute substantially to the scale enigma.

Finally, local proxies may be more variable than model simulations due to differences in spatial scales between local proxy records and a typical grid-box size in climate models (~100 km). However, for surface temperature, the spatial decorrelation length on interannual scales is usually larger than the size of a model grid box, which limits the impact of differences in spatial scale[54]. Moreover, as the spatial coherence of temperature variability increases between interannual and supradecadal timescales[55], discrepancies are expected to diminish towards longer timescales.

The described methodological differences between global and local reconstructions complicate the assessment of systematic biases. Nevertheless, current evidence suggests that neither an overestimation of local variability nor an underestimation of global variability from reconstructions can solely explain the diagnosed scale enigma. Therefore, we argue that model deficiencies substantially contribute to the enigma.

## Consequences for the spatial structure of natural climate variability

As a possible explanation for the scale enigma, we hypothesize that while the current generation of climate models can accurately simulate global-scale temperature variability, they are too stable at smaller (local) scales. If correct, this hypothesis would imply that climate models need more spatially independent modes of supradecadal variability.

Global and local variability are not independent, they are connected by the spatial correlation structure of the temperature field. Stronger spatial correlations lead to smaller differences between regional and global temperature variability and vice versa[56]. Comparing the spectra of simulated local and global variability suggests that the ratio between both quantities decreases with longer timescales (Fig. 3a, grey shaded area). This is associated with an increase in the typical spatial extent of temperature anomalies on longer timescales[56] and is supported by instrumental observations on subdecadal scales[12,55]. The relationship between temporal and spatial scales can also be visualized by subtracting two mean temperature fields from successive time steps from one another (for example, from two adjacent years (Fig. 3b), from adjacent decades (Fig. 3c) or from even longer periods (Fig. 3d)). In the CMIP5/6 model simulations, the local and global variability magnitudes converge as timescales increase (grey shading in Fig. 3a), which is associated with a continuously increasing spatial scale of the temperature fluctuations (Fig. 3b–d).

This increase in the spatial extent of temperature anomalies on supradecadal timescales might be an artefact introduced by the climate models, as the reconstructed local temperature variability on these timescales is larger than the simulated variability (Figs. 1 and 2). By adding published estimates of the local variability spectrum from PAGES 2k reconstructions[19] and marine and lake sediment cores[4,14] (Fig. 3a, blue dashed line; Methods) to the model spectrum, we tentatively deduce that the observed supradecadal local variability increases in parallel with the global variability. This would lead to a constant ratio between local and global variability (vertical dashed shading) and imply that, in contrast to the model results, the spatial scale of supradecadal temperature fluctuations might not increase after all. This prompts us to speculate that the true temperature anomalies between centuries might be more similar in structure and amplitude to the simulated temperature anomalies between decades (Fig. 3c), rather than the simulated anomalies between centuries (Fig. 3d).

This implied invariance in the spatial extent of temperature anomalies on supradecadal timescales according to the reconstructions, contrasting the further increase suggested by the models, may result from a different balance between two competing processes: (1) those creating local climate variability and (2) those causing a timescale-dependent suppression of local variability.

It has been proposed that the amplitude of natural forcing or the response to that forcing could be underestimated or be too spatially homogeneous due to shortcomings in the dynamical response in models[57]. Missing components of the climate system such as interactive ice sheets and dynamic vegetation[58–62] might further cause a lack of regional variability, especially when the mechanisms causing the variability are linked to timescale-invariant spatial structures such as the

https://doi.org/10.1038/s41561-023-01299-9

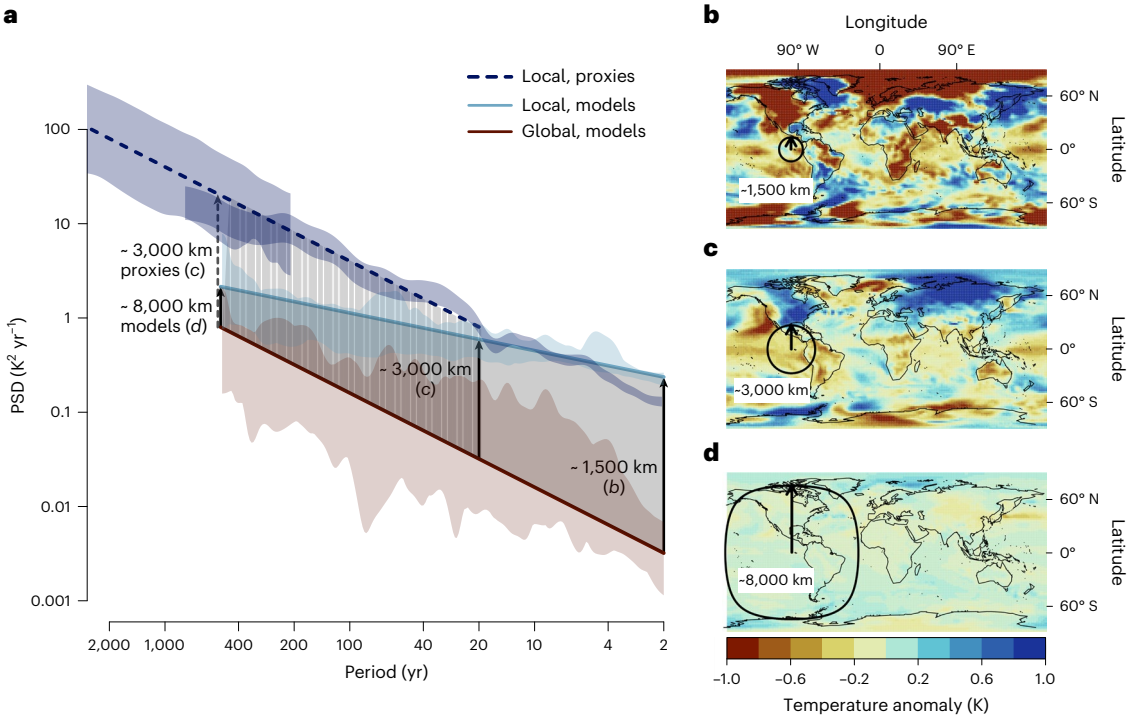



**Fig. 3 | Local versus global temperature variability and its implied spatial scale. a**, Global (brown) and mean local (light blue) PSD estimates from CMIP5/6 and proxy-based mean local spectra (dark blue) (refs. 4,14,19 and Methods). Lines show slopes PSD $\approx f^{-1}$ (red and dashed dark blue) and PSD $\approx f^{-0.4}$ (light blue) for frequencies $f$; the shading shows the range of all estimates. The ratio between local and global spectra (represented by a distance in this plot as due to the logarithmic $y$-axis) relates to the typical sizes of temperature anomalies[56] (vertical arrows). For models, the ratio decreases with increasing timescales (grey shading), whereas according to proxies it remains constant on supradecadal timescales (vertical dashed grey shading). **b**–**d**, Typical model temperature anomaly fields for 2 yr (**b**), 20 yr (**c**) and 400 yr (**d**) timescales. Circles show the typical spatial extents of temperature anomalies[55]. We hypothesize that real-world temperature anomalies on supradecadal timescales might be closer to those shown in **c** than those shown in **d**.

topography or land–sea boundaries. On the other hand, horizontal diffusion[63] reduces regional variability at low frequencies. Hence, overly weak or misrepresented ocean eddies[64] and inadequacies in the representation and spatiotemporal propagation of sub-grid-scale processes in current climate models could artificially dampen supradecadal temperature variabilities at regional scales. One notable limitation of the current CMIP protocols for past simulations that could contribute to the local variability deficit of simulations is the lack of knowledge regarding the dynamic initial state of the ocean. The Earth system has a long memory for past forcings[65] and the supradecadal variability has been suggested to echo slowly varying boundary conditions, mainly orbital forcing[10]. Meanwhile, simulations of the past millennium are generally initialized in a state of quasi-equilibrium with the starting boundary conditions.

Our hypothesis of a stronger persistence of regional temperature variability is consistent with other observational and proxy evidence. A lack of temporal persistence resulting in too much fast variation relative to to slow variation of regional temperatures in climate models[66,67] and a local response to forcing that is too weak were proposed as explanations for the so-called 'signal-to-noise paradox'[66]. This denotes the perplexing finding that models predict the observed (regional) temperature variations better than they predict ensemble forecast members. Finally, proxy-based reconstructions of temperature anomalies for specific periods in the past often exhibit more complex spatial patterns than the corresponding fields from simulations[68–70].

## Implications for climate projections and attribution efforts

Misrepresented variability would decrease the skill of policy-relevant climate and extreme event projections[71,72], impact assessments[73,74] and attribution studies[75]. The ability of forced climate model simulations to reproduce the magnitude and structure of global temperature variability (Fig. 2) strengthens the confidence in the simulated response to external forcing[9,76] in future climate projections. In contrast, biases in the spatial structure of variability would affect regional climate projections at supradecadal timescales. Likewise, projections might miss abrupt climate transitions as many climate models have been shown to be too stable to accurately simulate such transitions in the past[77]. However, recent progress in Earth system models[60,78] regarding the representation of these (in)stabilities of the climate system is promising.

Long-term variability modulates the occurrence of extreme weather and climate events[72,79], and stronger temporal persistence can increase the frequency of occurrence of such extremes (Fig. 4). As a result, current model-based loss and damage estimates from extreme events and risk projections may be underestimations. This may also affect short-term projections as the return periods of extreme events might be misrepresented (Fig. 4c). Finally, our reliance on instrumental observations can bias the distribution estimates of climate variables (Fig. 4b) as this period might be too short to sample the full range of supradecadal variability. Stronger persistence of climate variations would increase the probability of extreme conditions (Fig. 4a), including very rare and intense events[80]. Compound events may also occur more often, intensifying the vulnerability of ecosystems and societies, and the risk of sustained impacts such as agricultural losses and wildfires[81]. Socio-economic long-term planning for climate change requires a better understanding of this linkage between variability and extreme events[82].

Attribution studies provide vital insights into the impacts of anthropogenic greenhouse gas emissions on current and future climate change, including extreme climate and weather events. The reliability

https://doi.org/10.1038/s41561-023-01299-9

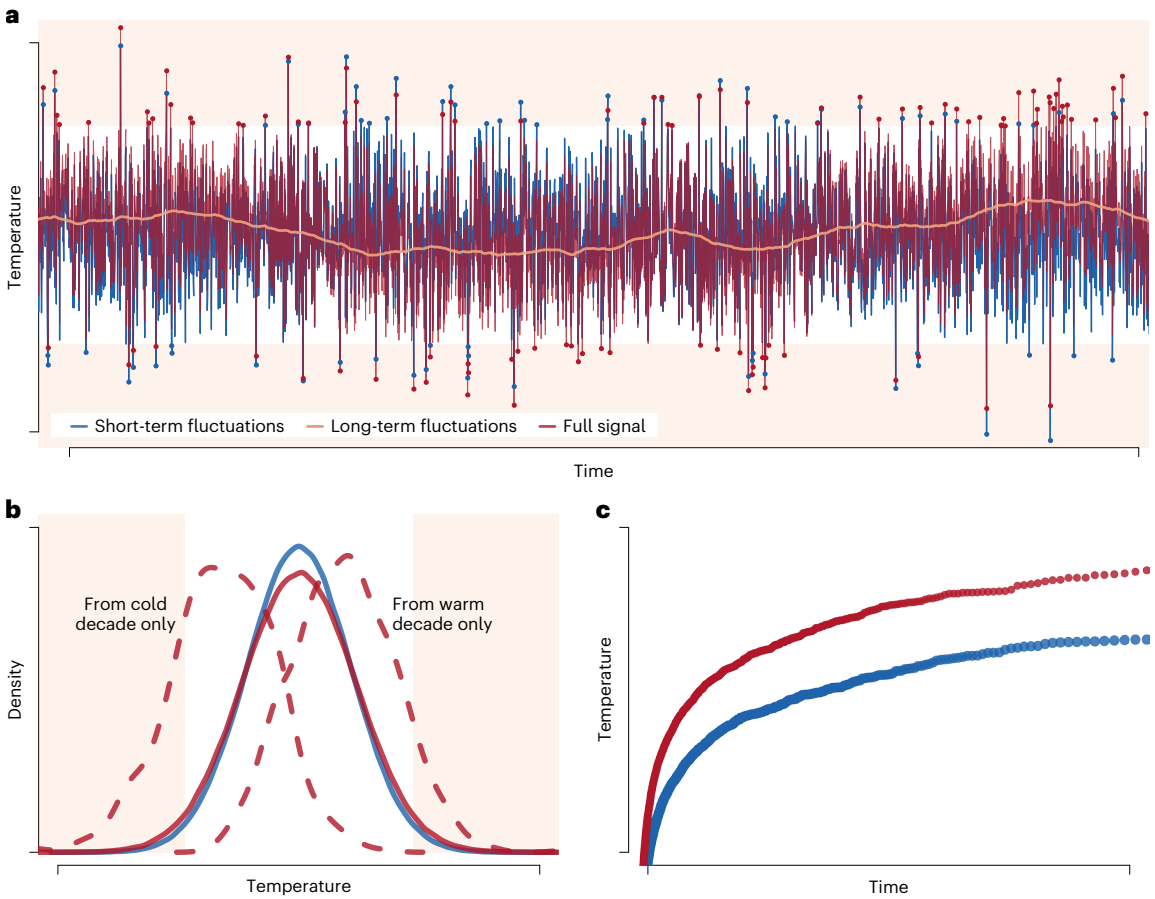



**Fig. 4 | Slow variability increases the magnitude of extreme events and decreases their return time in a conceptual time series. a**, Conceptual time series of a white noise process representing short-term (that is, weather; light blue) and an autocorrelated process representing long-term fluctuations (that is, decadal-to-centennial variability; orange). The sum of both (red) exemplifies a typical climate time series of 50 years (Methods). Dots indicate extremes more than two standard deviations from the mean (shaded regions). **b**, Probability density function of short-term and full signals (solid lines) and of the full signal estimated from an exceptionally cold and warm decade (dashed lines). **c**, Return levels from the short-term and full signals.

of these assessments depends on accurate representations of the statistical properties of the climate and its decomposition into anthropogenic and natural contributions[83,84], which makes these assessments susceptible to misrepresentations of variability in models[75]. Indeed, extreme event attribution often assumes that decadal and longer internal variability plays only a minor role in the occurrence of extremes[85].

Discrepancies between modelled and observed changes in climatic parameters complicate regional attribution[86] as most attribution studies focus precisely on those local to regional spatial scales at which the supradecadal variability is probably underestimated in models. Thus, if strong and persistent local variability is missing in models, tests for the significance of anthropogenic effects are likely to be biased positive. These arguments highlight the urgent need for an improved understanding of variability and its spatial structure, including the relationship between local and global variability estimates. We could then leverage this new understanding to validate and improve the capabilities of climate models regarding regional projections, for example, through emergent constraints[87], downscaling or bias-correction techniques[88] targeting variability. Incorporating stochastic natural forcing in climate model simulations of future scenarios is another important step in climate modelling for realistic projections of future climate variability[89,90].

Regional projections, risk assessments of extremes and attribution all depend on reliable simulations of local and regional variability. Without an improved representation of variability in climate models, the range of future conditions might not be fully covered. Thus, policies that are based on these estimates will not address the actual required mitigation and adaptation strategies. Overall, underestimation of the full temperature range can lead to a false sense of security (during a decade on the lower end of the range, for example) (Fig. 4), as well as to decades of unexpectedly high temperatures[1] for which policymakers and societies are unprepared.

## Closing the existing knowledge gap

To resolve the enigma of global climate models and proxy-based climate reconstructions being in agreement regarding the global average continuum of temperature variability but not at smaller (local to regional) scales, we have argued that regional-scale fluctuations in climate models should be larger and persist for longer. As a result, regional temperatures in current models probably seem too stable on supradecadal timescales and internal variability constitutes only a small fraction of their projection uncertainties, even at regional scales[91]. Additional efforts are required to ensure that natural (forced and internal) variability is simulated accurately to better evaluate its contribution to regional projection uncertainty. This includes investigating long model simulations with transient boundary conditions and simulations with additional coupled Earth system components in terms of their simulated supradecadal variability.

Validation of climate models with respect to palaeoclimate data usually focuses on their ability to simulate the mean state in a given time period such as the Last Glacial Maximum or slow changes, but rarely with respect to variability across timescales; this requires

both a paradigm shift in our approach to validating models and a new data product. In the 1970s, Climate: Long range Investigation, Mapping, and Prediction[92] became a landmark project, providing the first spatially comprehensive estimates of climate change and mapping the climate and ecology of the Last Glacial Maximum. An undertaking of similar magnitude concerning natural variability is now necessary and achievable. It requires the participation of the palaeoclimate, proxy-development and statistics communities. While we already have a large number of palaeoclimate reconstructions that cover several continents and oceanic basins, as well as timescales that range from annual to millennia, this dataset needs to be extended to improve replicability[93] and reduce the number and size of gaps in spatial, temporal and timescale coverage. While this may seem to be a considerable challenge, recent advances in analytical techniques such as mass spectral imaging[94] or laser spectroscopy[95] can be exploited to help achieve this, as these methods allow faster, more efficient and higher-resolution processing of proxy data. We need to combine community-driven synthesis and production of proxy datasets with a process-based understanding of the recording of climate signals in different types of archive[30] based on more site-level monitoring data. This will facilitate the separation of the climate signal from non-climatic noise, the separation of temperature and hydroclimate variability and the identification of the timescales at which the climate signal dominates variability. We also need to further develop and optimize statistical methods to combine the diverse information into reliable and unbiased variability estimates. This effort would facilitate spatially comprehensive estimates of variability as a function of timescale, ideally both for surface temperature and hydroclimatic variables.

Such advances are crucial for benchmarking and improving the representation of variability in climate models and the quantification of future regional climate risks linked to natural variability. Making use of distinct spatial fingerprints of the different mechanisms producing supradecadal variability would help identify the origins of the currently missing local supradecadal variability in models. Furthermore, these advances can provide the basis for characterizing the spatiotemporal covariance structure of climate variability. This in turn will improve the skill of data assimilation[96] and climate field reconstruction methods, permitting us to test our hypothesis for the scale enigma.

Finally, variability is a critical criterion of model development that should be evaluated in validating models and as an emergent property. This would allow us to investigate a broader range of possible model sensitivities and predictions of climate variability at various spatiotemporal scales. The proposed measures are vital for reliable estimates of variability, which in turn advance more robust regional projections, assessments of model-independent constraints and attribution and impact studies. The improved understanding of climate variability would thus contribute to better climate policy choices and mitigation and adaptation efforts.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41561-023-01299-9.

## References


1. IPCC *Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) (Cambridge Univ. Press, 2021).
2. Degroot, D. et al. Towards a rigorous understanding of societal responses to climate change. *Nature* **591**, 539–550 (2021).
3. Deser, C., Phillips, A., Bourdette, V. & Teng, H. Uncertainty in climate change projections: the role of internal variability. *Clim. Dynam.* **38**, 527–546 (2012).
4. Hébert, R., Herzschuh, U. & Laepple, T. Millennial-scale climate variability over land overprinted by ocean temperature fluctuations. *Nat. Geosci.* **15**, 899–905 (2022).
5. Maher, N., Lehner, F. & Marotzke, J. Quantifying the role of internal variability in the temperature we expect to observe in the coming decades. *Environ. Res. Lett.* **15**, 054014 (2020).
6. Hourdin, F. et al. The art and science of climate model tuning. *Bull. Am. Meteorol. Soc.* **98**, 589–602 (2017).
7. McKinnon, K. A. & Deser, C. Internal variability and regional climate trends in an observational large ensemble. *J. Clim.* **31**, 6783–6802 (2018).
8. Fredriksen, H.-B. & Rypdal, K. Spectral characteristics of instrumental and climate model surface temperatures. *J. Clim.* **29**, 1253–1268 (2016).
9. Crowley, T. J. Causes of climate change over the past 1000 years. *Science* **289**, 270–277 (2000).
10. Zhu, F. et al. Climate models can correctly simulate the continuum of global-average temperature variability. *Proc. Natl Acad. Sci. USA* **116**, 8728–8733 (2019).
11. Fernández-Donado, L. et al. Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium. *Clim. Past* **9**, 393–421 (2013).
12. Laepple, T. & Huybers, P. Global and regional variability in marine surface temperatures. *Geophys. Res. Lett.* **41**, 2528–2534 (2014).
13. Parsons, L. A. et al. Temperature and precipitation variance in CMIP5 simulations and paleoclimate records of the last millennium. *J. Clim.* **30**, 8885–8912 (2017).
14. Laepple, T. & Huybers, P. Ocean surface temperature variability: large model–data differences at decadal and longer periods. *Proc. Natl Acad. Sci. USA* **111**, 16682–16687 (2014).
15. Rehfeld, K., Münch, T., Ho, S. L. & Laepple, T. Global patterns of declining temperature variability from the Last Glacial Maximum to the Holocene. *Nature* **554**, 356–359 (2018).
16. Neukom, R. et al. Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nat. Geosci.* **12**, 643–649 (2019).
17. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
18. Black, D. E. et al. An 8-century tropical Atlantic SST record from the Cariaco Basin: baseline variability, twentieth-century warming, and Atlantic hurricane frequency. *Paleoceanogr. Palaeoclimatol.* **22**, PA4204 (2007).
19. Ellerhoff, B. & Rehfeld, K. 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).
20. Askjær, T. G. et al. Multi-centennial Holocene climate variability in proxy records and transient model simulations. *Quat. Sci. Rev.* **296**, 107801 (2022).
21. Cheung, A. H. et al. Comparison of low-frequency internal climate variability in CMIP5 models and observations. *J. Clim.* **30**, 4763–4776 (2017).
22. Bothe, O., Jungclaus, J. H. & Zanchettin, D. Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble. *Clim. Past* **9**, 2471–2487 (2013).
23. Collins, M., Osborn, T. J., Tett, S. F. B., Briffa, K. R. & Schweingruber, F. H. A comparison of the variability of a climate model with paleotemperature estimates from a network of tree-ring densities. *J. Clim.* **15**, 1497–1515 (2002).
24. Ault, T. R., Deser, C., Newman, M. & Emile-Geay, J. Characterizing decadal to centennial variability in the equatorial Pacific during the last millennium. *Geophys. Res. Lett.* **40**, 3450–3456 (2013).

25. Bühler, J. C. et al. Comparison of the oxygen isotope signatures in speleothem records and iHadCM3 model simulations for the last millennium. *Clim. Past* **17**, 985–1004 (2021).
26. Zorita, E. et al. European temperature records of the past five centuries based on documentary/instrumental information compared to climate simulations. *Climatic Change* **101**, 143–168 (2010).
27. Dee, S. G. et al. Improved spectral comparisons of paleoclimate models and observations via proxy system modeling: Implications for multi-decadal variability. *Earth Planet. Sci. Lett.* **476**, 34–46 (2017).
28. Franke, J., Frank, D., Raible, C. C., Esper, J. & Brönnimann, S. Spectral biases in tree-ring climate proxies. *Nat. Clim. Change* **3**, 360–364 (2013).
29. PAGES 2k-PMIP3 group. Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Clim. Past* **11**, 1673–1699 (2015).
30. Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M. & Anchukaitis, K. J. Applications of proxy system modeling in high resolution paleoclimatology. *Quat. Sci. Rev.* **76**, 16–28 (2013).
31. Anchukaitis, K. J. & Smerdon, J. E. Progress and uncertainties in global and hemispheric temperature reconstructions of the Common Era. *Quat. Sci. Rev.* **286**, 107537 (2022).
32. Esper, J., Frank, D. C. & Wilson, R. J. S. Climate reconstructions: low-frequency ambition and high-frequency ratification. *Eos* **85**, 113–120 (2004).
33. Kunz, T., Dolman, A. M. & Laepple, T. A spectral approach to estimating the timescale-dependent uncertainty of paleoclimate records – part 1: theoretical concept. *Clim. Past* **16**, 1469–1492 (2020).
34. Christiansen, B. & Ljungqvist, F. C. Challenges and perspectives for large-scale temperature reconstructions of the past two millennia. *Rev. Geophys.* **55**, 40–96 (2017).
35. Osborn, T. J. CLIMATE: the real color of climate change? *Science* **306**, 621–622 (2004).
36. Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A. & Funkhouser, G. The 'segment length curse' in long tree-ring chronology development for palaeoclimatic studies. *Holocene* **5**, 229–237 (1995).
37. Tingley, M. P. & Huybers, P. A Bayesian algorithm for reconstructing climate anomalies in space and time. part i: development and applications to paleoclimate reconstruction problems. *J. Clim.* **23**, 2759–2781 (2009).
38. Moberg, A., Mohammad, R. & Mauritsen, T. Analysis of the Moberg et al. (2005) hemispheric temperature reconstruction. *Clim. Dynam.* **31**, 957–971 (2008).
39. Trouet, V. et al. A 1500-year reconstruction of annual mean temperature for temperate North America on decadal-to-multidecadal time scales. *Environ. Res. Lett.* **8**, 024008 (2013).
40. Kim, S.-T. & O'Neil, J. R. Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates. *Geochim. Cosmochim. Acta* **61**, 3461–3475 (1997).
41. Werner, M., Mikolajewicz, U., Heimann, M. & Hoffmann, G. Borehole versus isotope temperatures on Greenland: seasonality does matter. *Geophys. Res. Lett.* **27**, 723–726 (2000).
42. Müller, P. J., Kirst, G., Ruhland, G., Von Storch, I. & Rosell-Melé, A. Calibration of the alkenone paleotemperature index $U^{K'}_{37}$ based on core-tops from the eastern South Atlantic and the global ocean (60°N-60°S). *Geochim. Cosmochim. Acta* **62**, 1757–1772 (1998).
43. Laepple, T. et al. On the similarity and apparent cycles of isotopic variations in East Antarctic snow pits. *Cryosphere* **12**, 169–187 (2018).
44. Zuhr, A. M. et al. Age-heterogeneity in marine sediments revealed by three-dimensional high-resolution radiocarbon measurements. *Front. Earth Sci.* https://doi.org/10.3389/feart.2022.871902 (2022).
45. Peeters, F. J. C., Brummer, G.-J. A. & Ganssen, G. The effect of upwelling on the distribution and stable isotope composition of *Globigerina bulloides* and *Globigerinoides ruber* (planktic foraminifera) in modern surface waters of the NW Arabian Sea. *Glob. Planet. Change* **34**, 269–291 (2002).
46. Berger, W. H. & Heath, G. R. Vertical mixing in pelagic sediments. *J. Mar. Res.* **26**, 134–143 (1968).
47. Johnsen, S. J. in *Isotopes and Impurities in Snow and Ice* Publication No. 118, 210–219 (IAHS-AISH, 1977).
48. Webb, T. Is vegetation in equilibrium with climate? How to interpret late-Quaternary pollen data. *Vegetatio* **67**, 75–91 (1986).
49. Mix, A. in *North America and Adjacent Oceans During the Last Deglaciation* Vol. K-3, 111–135 (Geological Society of America, 1987).
50. Laepple, T. & Huybers, P. Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability. *Earth Planet. Sci. Lett.* **375**, 418–429 (2013).
51. Dee, S. et al. PRYSM: an open-source framework for proxy system modeling, with applications to oxygen-isotope systems. *J. Adv. Model. Earth Syst.* **7**, 1220–1247 (2015).
52. Rhines, A. & Huybers, P. Estimation of spectral power laws in time uncertain series of data with application to the Greenland Ice Sheet Project 2 $\delta^{18}O$ record. *J. Geophys. Res. Atmos.* **116**, D01103 (2011).
53. Sigl, M. et al. Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature* **523**, 543–549 (2015).
54. North, G. R., Wang, J. & Genton, M. G. Correlation models for temperature fields. *J. Clim.* **24**, 5850–5862 (2011).
55. Jones, P. D., Osborn, T. J. & Briffa, K. R. Estimating sampling errors in large-scale temperature averages. *J. Clim.* **10**, 2548–2568 (1997).
56. Kunz, T. & Laepple, T. Frequency-dependent estimation of effective spatial degrees of freedom. *J. Clim.* **34**, 7373–7388 (2021).
57. Shindell, D. T., Schmidt, G. A., Mann, M. E., Rind, D. & Waple, A. Solar forcing of regional climate change during the Maunder Minimum. *Science* **294**, 2149–2152 (2001).
58. Bakker, P., Clark, P. U., Golledge, N. R., Schmittner, A. & Weber, M. E. Centennial-scale Holocene climate variations amplified by Antarctic Ice Sheet discharge. *Nature* **541**, 72–76 (2017).
59. Braconnot, P., Zhu, D., Marti, O. & Servonnat, J. Strengths and challenges for transient Mid- to Late Holocene simulations with dynamical vegetation. *Clim. Past* **15**, 997–1024 (2019).
60. Hopcroft, P. O. & Valdes, P. J. Paleoclimate-conditioning reveals a North Africa land–atmosphere tipping point. *Proc. Natl Acad. Sci. USA* **118**, e2108783118 (2021).
61. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449 (2008).
62. Laguë, M. M., Bonan, G. B. & Swann, A. L. S. Separating the impact of individual land surface properties on the terrestrial surface energy budget in both the coupled and uncoupled land–atmosphere system. *J. Clim.* **32**, 5725–5744 (2019).
63. Rypdal, K., Rypdal, M. & Fredriksen, H.-B. Spatiotemporal long-range persistence in Earth's temperature field: analysis of stochastic–diffusive energy balance models. *J. Clim.* **28**, 8379–8395 (2015).
64. Jüling, A., von der Heydt, A. & Dijkstra, H. A. Effects of strongly eddying oceans on multidecadal climate variability in the Community Earth System Model. *Ocean Sci.* **17**, 1251–1271 (2021).
65. Rypdal, M. & Rypdal, K. Long-memory effects in linear response models of Earth's temperature and implications for future global warming. *J. Clim.* **27**, 5240–5258 (2014).

66. Sevellec, F. & Drijfhout, S. S. The signal-to-noise paradox for interannual surface atmospheric temperature predictions. *Geophys. Res. Lett.* **46**, 9031–9041 (2019).
67. Strommen, K. & Palmer, T. N. Signal and noise in regime systems: a hypothesis on the predictability of the North Atlantic Oscillation. *Q. J. R. Meteorol. Soc.* **145**, 147–163 (2019).
68. Mann, M. E. et al. Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. *Science* **326**, 1256–1260 (2009).
69. Hargreaves, J. C., Annan, J. D., Ohgaito, R., Paul, A. & Abe-Ouchi, A. Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene. *Clim. Past* **9**, 811–823 (2013).
70. Weitzel, N., Hense, A. & Ohlwein, C. Combining a pollen and macrofossil synthesis with climate simulations for spatial reconstructions of European climate using Bayesian filtering. *Clim. Past* **15**, 1275–1301 (2019).
71. Blanusa, M. L., López-Zurita, C. J., & Rasp, S. Internal variability plays a dominant role in global climate projections of temperature and precipitation extremes. *Climate Dynamics* **61**, 1931–1945 (2023).
72. Ionita, M., Dima, M., Nagavciuc, V., Scholz, P. & Lohmann, G. Past megadroughts in central Europe were longer, more severe and less warm than modern droughts. *Commun. Earth Environ.* **2**, 61 (2021).
73. Calel, R., Chapman, S. C., Stainforth, D. A. & Watkins, N. W. Temperature variability implies greater economic damages from climate change. *Nat. Commun.* **11**, 5028 (2020).
74. Schwarzwald, K. & Lenssen, N. The importance of internal climate variability in climate impact projections. *Proc. Natl Acad. Sci. USA* **119**, e2208095119 (2022).
75. Harrington, L. J., Schleussner, C.-F. & Otto, F. E. L. Quantifying uncertainty in aggregated climate change risk assessments. *Nat. Commun.* **12**, 7140 (2021).
76. Hausfather, Z., Drake, H. F., Abbott, T. & Schmidt, G. A. Evaluating the performance of past climate model projections. *Geophys. Res. Lett.* **47**, e2019GL085378 (2020).
77. Valdes, P. Built for stability. *Nat. Geosci.* **4**, 414–416 (2011).
78. Klockmann, M., Mikolajewicz, U., Kleppin, H. & Marotzke, J. Coupling of the subpolar gyre and the overturning circulation during abrupt glacial climate transitions. *Geophys. Res. Lett.* **47**, e2020GL090361 (2020).
79. Czymzik, M., Muscheler, R. & Brauer, A. Solar modulation of flood frequency in central Europe during spring and summer on interannual to multi-centennial timescales. *Clim. Past* **12**, 799–805 (2016).
80. Yan, M. & Liu, J. Physical processes of cooling and mega-drought during the 4.2 ka BP event: results from TraCE-21ka simulations. *Clim. Past* **15**, 265–277 (2019).
81. Zscheischler, J. et al. A typology of compound weather and climate events. *Nat. Rev. Earth Environ.* **1**, 333–347 (2020).
82. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nat. Clim. Change* **2**, 775–779 (2012).
83. Hegerl, G. & Zwiers, F. Use of models in detection and attribution of climate change. *WIREs Clim. Change* **2**, 570–591 (2011).
84. Stott, P. A. et al. Observational constraints on past attributable warming and predictions of future global warming. *J. Clim.* **19**, 3055–3069 (2006).
85. Philip, S. et al. A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **6**, 177–203 (2020).
86. van Oldenborgh, G. J. et al. Pathways and pitfalls in extreme event attribution. *Climatic Change* **166**, 13 (2021).
87. Qasmi, S. & Ribes, A. Reducing uncertainty in local temperature projections. *Sci. Adv.* **8**, eabo6872 (2022).
88. Wu, Y. et al. Quantifying the uncertainty sources of future climate projections and narrowing uncertainties with bias correction techniques. *Earth Future* **10**, e2022EF002963 (2022).
89. Bethke, I. et al. Potential volcanic impacts on future climate variability. *Nat. Clim. Change* **7**, 799–805 (2017).
90. Ellerhoff, B. et al. Contrasting state-dependent effects of natural forcing on global and local climate variability. *Geophys. Res. Lett.* **49**, e2022GL098335 (2022).
91. Lehner, F. et al. Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dynam.* **11**, 491–508 (2020).
92. McIntyre, A. et al. *Seasonal Reconstructions of the Earth's Surface at the Last Glacial Maximum* (Geological Society of America, 1981).
93. Comboul, M., Emile-Geay, J., Hakim, G. J. & Evans, M. N. Paleoclimate sampling as a sensor placement problem. *J. Clim.* **28**, 7717–7740 (2015).
94. Wörmer, L. et al. Ultra-high-resolution paleoenvironmental records via direct laser-based analysis of lipid biomarkers in sediment core samples. *Proc. Natl Acad. Sci. USA* **111**, 15669–15674 (2014).
95. Barkan, E. & Luz, B. High precision measurements of $^{17}O/^{16}O$ and $^{18}O/^{16}O$ ratios in $H_2O$. *Rapid Commun. Mass Spectrom.* **19**, 3737–3742 (2005).
96. Amrhein, D. E., Hakim, G. J. & Parsons, L. A. Quantifying structural uncertainty in paleoclimate data assimilation with an application to the last millennium. *Geophys. Res. Lett.* **47**, e2020GL090485 (2020).
97. Ljungqvist, F. C. et al. Centennial-scale temperature change in last millennium simulations and proxy-based reconstructions. *J. Clim.* **32**, 2441–2482 (2019).



[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany. [2]MARUM–Center for Marine Environmental Sciences and Faculty of Geosciences, University of Bremen, Bremen, Germany. [3]Department of Geosciences, University of Tübingen, Tübingen, Germany. [4]Department of Physics, University of Tübingen, Tübingen, Germany. [5]Institut für Naturwissenschaften, Geographie und Technik, Pädagogische Hochschule Heidelberg, Heidelberg, Germany. [6]Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Spain. [7]Institute for Coastal Systems—Analysis and Modelling, Helmholtz-Zentrum Hereon, Geesthacht, Germany. [8]Barcelona Supercomputing Center (BSC), Barcelona, Spain. [9]Section Meteorology, Institute of Geosciences, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany. [10]Institute of Earth Surface Dynamics, Geopolis, University of Lausanne, Lausanne, Switzerland. [11]ARC Centre of Excellence in Australian Biodiversity and Heritage (CABAH), Canberra, Australian Capital Territory, Australia. [12]School of Culture, History and Language, The Australian National University, Canberra, Australian Capital Territory, Australia. [13]Present address: Department Greenhouse gas emission verification, Deutscher Wetterdienst, Offenbach, Germany.
✉e-mail: tlaepple@awi.de

## Methods

### Spectral estimates

The PSD estimates were calculated using the multitaper method[98] with three tapers and a time-bandwidth parameter $\omega = 2$. The PSD estimates were smoothed using a Gaussian kernel with a constant width of 0.03 on the (base 10) logarithmic timescale.

### Literature review and agreement

We included literature that covered the entire Holocene and applied a ranking system to indicate the level of (dis)agreement between climate reconstructions and models using the following five levels: disagreement, largely disagree, neutral, largely agree, and agreement. Levels were assigned on the basis of the respective study authors' original assessments of their model–data consistency that were based on various methods used for the model–data comparison (for example direct comparison, proxy modelling) and proxy calibration (for example temporal calibration, spatial calibration). We differentiated the results according to temporal and spatial scales wherever possible. Supplementary Table 1 contains the full review on which Fig. 2 is based, together with statements from the original papers that were the basis for assigning the levels of (dis)agreement. Extended Data Fig. 2 explicitly cites the individual studies.

### Local and global spectra and relationship to the spatial scale of temperature fluctuations

Global mean spectra (Figs. 1 and 3) were computed as the PSD of the global temperature time series. As an example of a local record, we chose the Cariaco Mg/Ca records as it is one of the highest-resolution marine records that spans most of the past millennium (1200–2000 CE). This record is not affected by bioturbation (due to its laminated sediment) and is based on a classical temperature proxy (Mg/Ca on planktic foraminifera). Mg/Ca is clearly attributed to temperature and we can use an independent calibration (Methods; see ref. 14), circumventing potential issues with temporal calibrations. Local mean spectra (Figs. 1 and 3 and Extended Data Fig. 1) were computed as area-weighted mean spectra of the local (grid box) temperature. For the model spectra, the grid boxes containing the proxy record were selected. The local mean spectrum from the PAGES 2k database (Fig. 3 and Extended Data Fig. 1) corresponds to the estimate in ref. 19. For this estimate, 101 proxy records were considered and selected according to their resolution, number of data points and coverage, as well as their maximum hiatus. The result is robust to the selection criteria[19]. In Fig. 3, the local spectrum from sedimentary records is also shown, covering timescales from 1/200–1/3000 years. The proxy spectra were created from the weighted average of local SST spectra from ref. 14 (arithmetic mean of Mg/Ca and Uk37 proxy-based spectra, where Uk37 is the unsaturation index of $C_{37}$ methyl alkenones[50]) and local terrestrial air temperature spectra from ref. 4 (2/3 × SST + 1/3 × terrestrial temperature). All three spectra show a similar scaling of PSD $\approx f^{-1}$ for frequency $f$ and are therefore parallel to the global mean temperature spectra from CMIP5/6 and the PAGES 2k reconstruction (Fig. 1). This implies that the spatial scale remains nearly constant on supradecadal timescales, independent of the absolute magnitude of local variability. The ratio of local to global variability can be interpreted as effective spatial degrees of freedom that can in turn be translated into a characteristic length scale (an effective correlation radius)[56] (see Fig. 3a and the circles and labelling in Fig. 3b–d). To visualize a typical amplitude and spatial scale of temperature anomalies at different timescales (Fig. 3b–d), we computed a Haar fluctuation at different timescales (that is, the difference between the average of the first and second half of a given time interval) from the MPI-ESM-P past1000 experiment. We show examples for one 2 yr (Fig. 3b), one 20 yr (Fig. 3c) and one 500 yr fluctuation (Fig. 3d); they correspond to, respectively, the average temperature for the year 1050 minus 1051, the years 1050–1059 minus 1060–1069 and the years 1050–1299 minus 1300–1549.

### Conceptual relation between slow climate variability and extremes

The time series of short-term variations in Fig. 4 was created as white noise, while the time series of long-term variations was generated as a low-pass filtered (cutoff 1/10 yr) stochastic process with PSD $\approx f^{-1}$. The variance ratio was 5:1. The parameters and time steps were chosen such that the short-term time series corresponds to fluctuations with a characteristic timescale of weeks and the long-term series to decadal fluctuations. In total, we generated a 10,000 yr time series to ensure that the distributions converged. The distribution and return levels were obtained empirically. We used the definition of the return level from the Weibull formula, which relates the inverse rank, $i$, of the sorted time series vector, $\mathbf{X}$, of length $N$ to the return period $R(x_i)$ via (Box 1):

$$\mathbf{X} = (x_1, \ldots x_N),\ x_i \geq x_{i+1}$$

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

## Data availability

The PAGES 2k palaeotemperature records (PAGES 2k v.2.0.0) are available at www.ncdc.noaa.gov/paleo/study/21171. The ensemble of global temperature reconstructions based on the PAGES2k16 data are available through the World Data Service (NOAA) Palaeoclimatology at https://www.ncdc.noaa.gov/paleo/study/26872 and via Figshare at https://doi.org/10.6084/m9.figshare.c.4507043. The pollen-based reconstructions are available via PANGEA at https://doi.pangaea.de/10.1594/PANGAEA.930512. The marine proxy data are available via PANGEA at https://doi.org/10.1594/PANGAEA.899489. The CMIP5 millennium simulations are available through the Earth System Grid Federation portal at https://esgf-data.dkrz.de. Source data are provided with this paper.

## References


98. Percival, D. B. & Walden, A. T. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques* (Cambridge Univ. Press, 1993).
99. Wu, T. et al. Global carbon budgets simulated by the Beijing Climate Center Climate System Model for the last century. *J. Geophys. Res. Atmos.* **118**, 4326–4347 (2013).
100. Gent, P. R. et al. The Community Climate System Model Version 4. *J. Clim.* **24**, 4973–4991 (2011).
101. Li, L. et al. The flexible global ocean-atmosphere-land system model, grid-point version 2: FGOALS-g2. *Adv. Atmos. Sci.* **30**, 543–560 (2013).
102. Schmidt, G. A. et al. Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.* **6**, 141–184 (2014).
103. Dufresne, J.-L. et al. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim. Dynam.* **40**, 2123–2165 (2013).
104. Roeckner, E. et al. Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *J. Clim.* **19**, 3771–3791 (2006).
105. Yukimoto, S. et al. A new global climate model of the Meteorological Research Institute: MRI-CGCM3 —model description and basic performance. *J. Meteorol. Soc. Jpn Ser. II* **90A**, 23–64 (2012).
106. Volodin, E. M. et al. Simulation of the modern climate using the INM-CM48 climate model. *Russ. J. Numer. Anal. Math. Model.* **33**, 367–374 (2018).
107. Hajima, T. et al. Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geosci. Model Dev.* **13**, 2197–2244 (2020).

108. Mauritsen, T. et al. Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing $CO_2$. *J. Adv. Model. Earth Syst.* **11**, 998–1038 (2019).
109. Yukimoto, S. et al. The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: description and basic evaluation of the physical component. *J. Meteorol. Soc. Jpn Ser. II* **97**, 931–965 (2019).

## Acknowledgements

This study was undertaken by members of CVAS and 2k Network, working groups of the Past Global Changes (PAGES) Global Research association. This is a contribution to the SPACE ERC, STACY and PALMOD projects. The SPACE ERC project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 716092). STACY has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project no. 395588486). This work has also been supported by the German Federal Ministry of Education and Research (BMBF), through the PalMod project (subprojects 01LP1926B (O.B.), 01LP1926D (M.C.) and 01LP1926C (B.E., P.S. and N.W.)) from the Research for Sustainability initiative (FONA). B.E. is supported by the Heinrich Böll Foundation. E.M.-C. was supported by the PARAMOUR project, funded by the Fonds de la Recherche Scientifique–FNRS and the FWO under the Excellence of Science (EOS) programme (grant no. O0100718F, EOS ID no. 30454083). A.H. was supported by a Legacy Grant from the Australian Research Council Centre of Excellence for Australian Biodiversity and Heritage. B.M. was supported by LINKA20102 and the Spanish Ministry of Science and Innovation project CEX2018-000794-S. The work originated from discussions at the CVAS working group of PAGES at a workshop at the Internationales Wissenschaftsforum Heidelberg, which was funded by a Hengstberger Prize. We thank N. Beech, C. Brierley, F. Gonzalez-Rouco and M. MacPartland for comments on earlier drafts of the manuscript. This manuscript uses data provided by the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP and PMIP. We thank the research groups for producing and kindly making their model outputs, measurements and palaeoclimate reconstructions available to us. Editorial assistance, in the form of language editing and correction, was provided by XpertScientific Editing and Consulting Services. We acknowledge support by the Open Access Publication Funds of Alfred-Wegener-Institut Helmholtz Zentrum für Polar- und Meeresforschung.

## Author contributions

T.L. led the synthesis and overall analysis; E.Z. coordinated the work. K.R. and T.L. initiated the project through the workshop. R.H. led the writing of the introductory paragraph. E.Z. led the writing on the remainder of the introduction regarding the conflicting evidence in the literature; E.Z., B.M. and P.S. led the literature review; A.H. and all authors contributed to it; B.E. and R.H. produced Fig. 1, P.S. produced Fig. 2. N.W. and T.L. led the writing of the section on reconstruction deficiencies; N.W., E.M.-C. & T.L. led the writing of the section on the consequences for the spatial structure; R.H & T.L. produced Fig. 3. E.Z and B.E. led the writing of the section on the implications for climate projections and attribution efforts; B.E. produced Fig. 4. R.H. led the writing of the concluding section. O.B. provided a critical check of the current literature O.B., N.W., B.E, R.H., E.Z., T.L., K.R., B.M., M.C., reviewed each section in detail. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

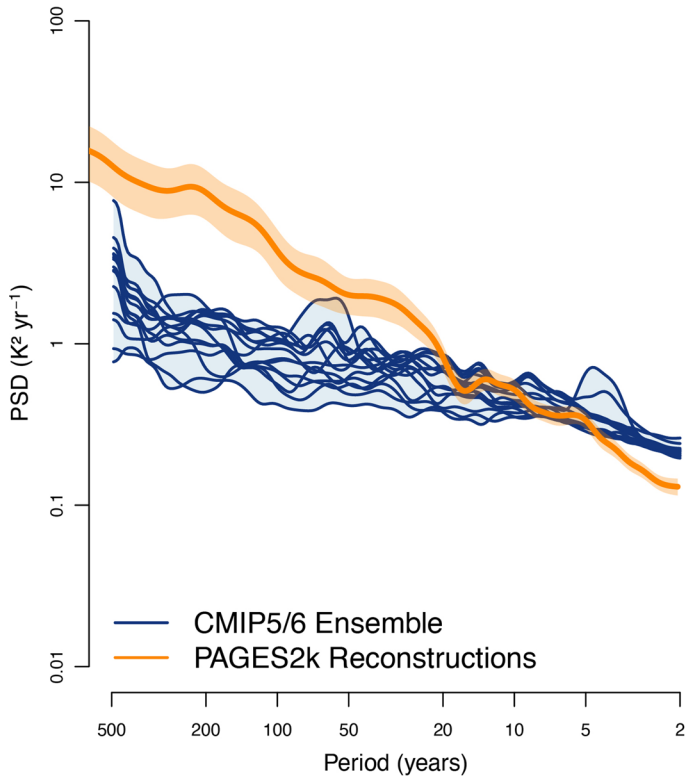
**Extended data** is available for this paper at https://doi.org/10.1038/s41561-023-01299-9.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41561-023-01299-9.

**Correspondence** should be addressed to T. Laepple.

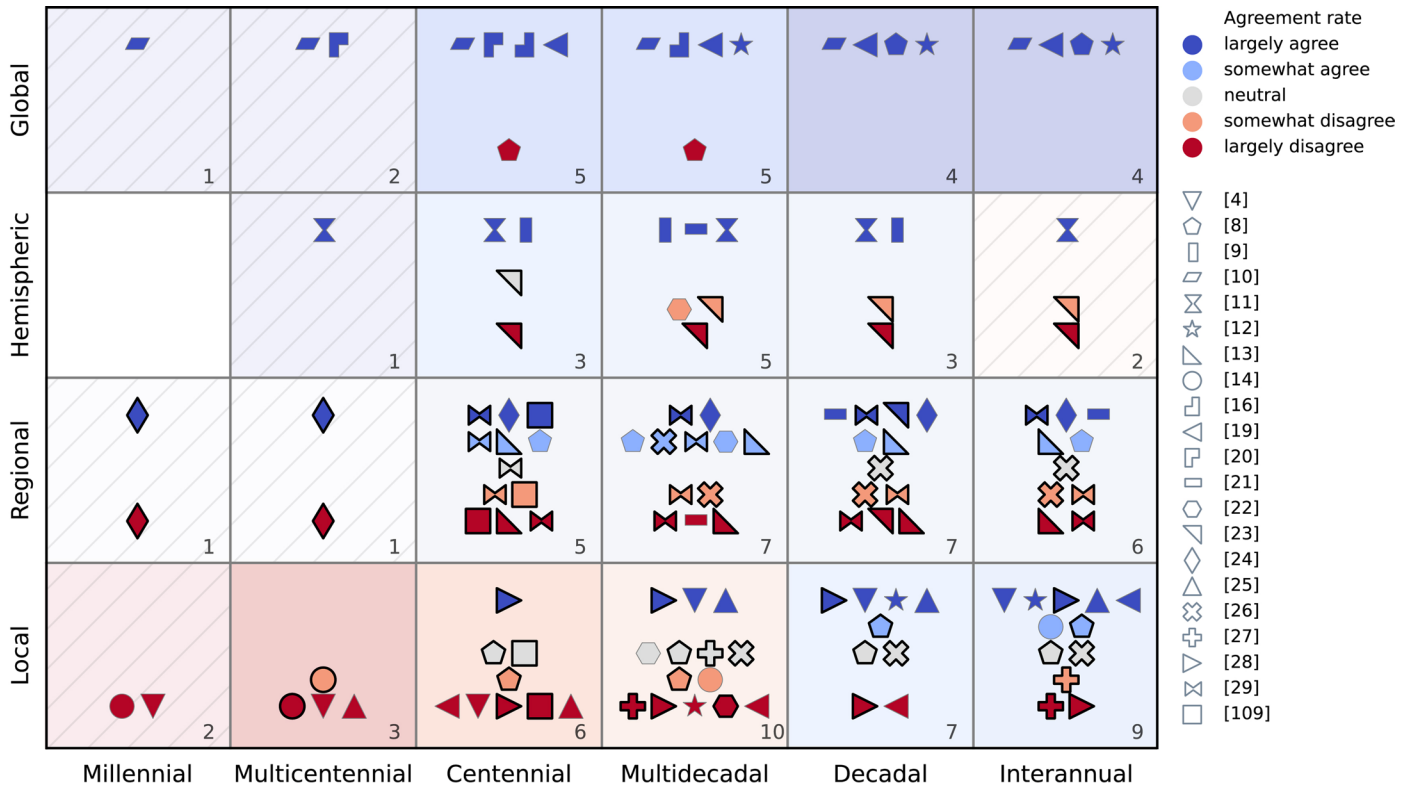
**Peer review information** *Nature Geoscience* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor(s): James Super, in collaboration with the *Nature Geoscience* team.

https://doi.org/10.1038/s41561-023-01299-9



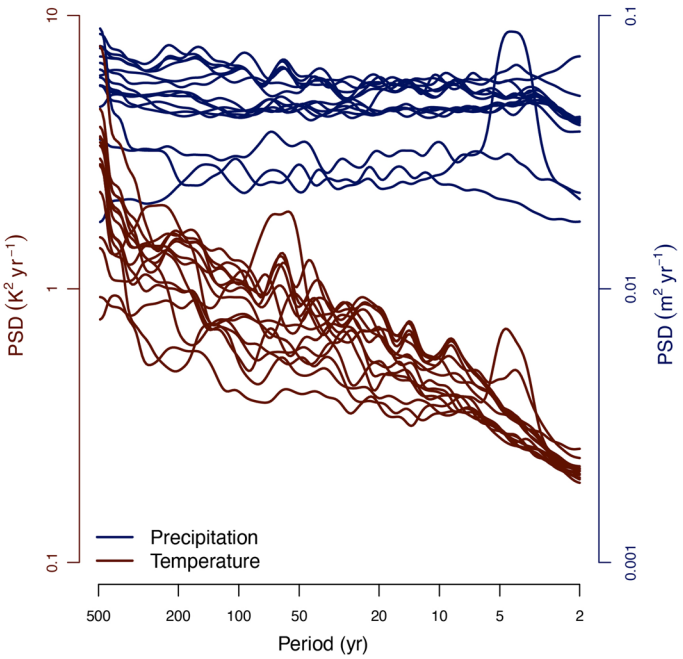

**Extended Data Fig. 1 | Spectrum of mean local simulated and reconstructed temperature variability.** As in Fig. 1d of the main text but for the mean local temperature spectrum from CMIP5/CMIP6 simulations and from PAGES2k temperature reconstructions (see Methods and Ref. 19). This shows that local variability reconstructed from instrumentally calibrated annually resolved records displays a similar model-data variability mismatch on supra-decadal time-scales as the example reconstruction (Cariaco) shown in Fig. 1d, or other marine or terrestrial records[4,14].

https://doi.org/10.1038/s41561-023-01299-9




**Extended Data Fig. 2 | Overview of model-data (dis)agreement in Holocene temperature variability in the literature with explicit references.** As in Fig. 2 of the main text, model-data agreement is grouped according to temporal (x-axis) and spatial scale (y-axis). Each symbol represents a specific study (refs. 4,8–14,16,19–2997) and the color-code indicates strength of (dis)agreement. Multiple occurrences in one box can happen when differing results are reported that is depending on reconstruction method or proxy type. Such cases are highlighted with a black border. The number at the bottom right of each box is the number of distinct studies in this box. Dashing of a box indicates only one or two studies for this spatio-temporal scale. Further details can be found in the Methods section.

https://doi.org/10.1038/s41561-023-01299-9




**Extended Data Fig. 3 | Local precipitation and local temperature variability show a different temporal scaling.** Local mean spectral estimates from CMIP5/6 precipitation (dark blue) and temperature (brown). Across all models, local precipitation variability shows a flatter (more white) scaling than local temperature variability. This implies that the mismatch between simulated and reconstructed local supra-decadal variability would increase, if the proxies would represent a mix of precipitation and temperature (calibrated to temperature units), as the difference in scaling between proxies and simulated precipitation is even larger than between proxies and simulated temperature.

https://doi.org/10.1038/s41561-023-01299-9

**Extended Data Table 1 | CMIP5/CMIP6 model experiments. The model experiments were used in Figs. 1 and 3 (refs. 99–109)**

| Model | Country | Atm. grid spacing (lat x lon) | Experiment | Runs | Generation | Reference |
|---|---|---|---|---|---|---|
| BCC-CSM1.1 | China | 2.8° x 2.8° | past1000 | r1i1p1 | CMIP5 | Ref[98] |
| CCSM4 | USA | 0.9° x 1.25° | past1000 | r1i1p1 | CMIP5 | Ref[99] |
| FGOALS-gl | China | 2.8° x 2.8° | past1000 | r1i1p1 | CMIP5 | Ref[100] |
| GISS-E2-R | USA | 2° x 2.5° | past1000 | r1i1p121<br>r1i1p122<br>r1i1p123<br>r1i1p124<br>r1i1p125<br>r1i1p126<br>r1i1p127<br>r1i1p128<br>r1i1p1221 | CMIP5 | Ref[101] |
| IPSL-CM5A-LR | France | 1.9° x 3.8° | past1000 | r1i1p1 | CMIP5 | Ref[102] |
| MPI-ESM | Germany | 1.875° x 1.875° | past1000 | r1i1p1 | CMIP5 | Ref[103] |
| MRI-CGCM3 | Japan | 1.125° x 1.125° | past1000 | r1i1p1 | CMIP5 | Ref[104] |
| INM-CM4-8 | Russia | 2° x 1.15° | past1000 | r1i1p1f1 | CMIP6 | Ref[105] |
| MIROC-ES2L | Japan | 2.8° x 2.8° | past1000 | r1i1p1f2 | CMIP6 | Ref[106] |
| MPI-ESM1-2-LR | Germany | 1.875° x 1.875° | past2k | r1i1p1f1 | CMIP6 | Ref[107] |
| MRI-ESM2-0 | Japan | 1.125° x 1.125° | past1000 | r1i1p1f1 | CMIP6 | Ref[108] |