



Quantifying structural and proxy composition uncertainty in Common Era climate field reconstructions using online paleo data assimilation

Beyrem Jebri, Myriam Khodri

5 LOCEAN/IPSL, Sorbonne Université, CNRS, Paris, 75005, France

Correspondence to: Beyrem Jebri (beyrem.jebri@locean.ipsl.fr) and Myriam Khodri (myriam.khodri@locean.ipsl.fr)

Abstract. Paleodata assimilation combines proxy observations with climate model information to produce spatially complete reconstructions of past climate, but uncertainty is often represented only by posterior ensemble spread within a single reconstruction setup. Here, we generate annually resolved global reconstructions of near surface air temperature, sea surface temperature, precipitation, evaporation and sea level pressure over the 1-2012AD Common Era using an online Sequential Importance Resampling particle filter coupled with Linear Inverse Models trained from CMIP5 and CMIP6 simulations. We quantify uncertainty with an 80 member factorial ensemble varying proxy system model calibration, model prior and bias correction, and a 50 member ensemble testing sensitivity to tree ring and coral proxy inclusion. We distinguish posterior internal uncertainty, structural uncertainty and proxy composition uncertainty. The reconstruction agrees best with instrumental products for large scale temperature and sea surface temperature fields, while hydroclimate and circulation variables show more heterogeneous skill. Structural uncertainty dominates the amplitude of secular global mean temperature and North Atlantic sea surface temperature variability, mainly through proxy system model calibration, whereas prior choice and bias correction become more important for tropical and circulation indices. The uncertainty framework identifies several coherent reconstructed features: robust industrial era global mean warming, a pre industrial warm optimum around 600-700 CE following colder conditions around 400-600 CE, a low frequency North Atlantic cooling toward the late Little Ice Age, and selected tropical Pacific anomalies around 950, 1200 and 1400 CE. Recent tropical Pacific changes include a positive Niño3.4 tendency and a strong negative Pacific Walker circulation excursion near 2000 CE, the latter suggesting a potentially unusual circulation state relative to the reconstructed Common Era background. Tree rings and corals provide complementary constraints, showing that proxy composition affects not only uncertainty width but also the interpretation of reconstructed climate features.

1 Introduction

The Common Era (CE; the last two millennia) offers one of the few intervals over which annually to decadal resolved proxy evidence, external forcing estimates and transient model simulations can be brought together to evaluate climate



30 variability beyond the instrumental record. This is essential because the instrumental period samples only a narrow set of
 ocean atmosphere states and is too short to characterize the persistence and amplitude of multidecadal to centennial
 variability, especially at regional scales (Hawkins et al., 2016; PAGES Hydro2k Consortium, 2017; Laepple et al., 2023).
 The CE is therefore not only a context for recent warming; it is a test bed for identifying dynamical states and forced
 disturbances that are absent or poorly sampled in modern observations. Existing CE reconstructions have established broad
 35 temperature features and increasingly provide spatially complete hydroclimate and circulation fields. Their differences,
 however, show that the central question is no longer simply whether a feature appears in one product, but whether its
 interpretation survives changes in method, prior covariance and proxy network composition. This issue is particularly acute
 for tropical Pacific and circulation indices, where model data disagreement, teleconnection nonstationarity and sparse
 tropical archives can make product comparison alone insufficient (Steiger et al., 2018; Falster et al., 2023; Luo et al., 2024).
 40 Paleoclimate data assimilation (PDA) is well suited to this problem because it combines proxy observations with model
 based covariance structures to reconstruct multivariate climate states. Yet PDA also makes uncertainty interpretation more
 demanding. Posterior particle spread within one configuration measures only the range of states compatible with a fixed
 prior, PSM and error model. It cannot determine whether the same feature would persist under alternative covariance
 structures, calibration choices or network compositions. This distinction is central because covariance uncertainty can
 45 strongly affect reconstructed fields in regions that are poorly constrained directly by proxies (Amrhein et al., 2020). Recent
 work on temporal comparisons in PDA further shows that uncertainty plumes cannot always be compared as if they were
 ensembles of physically connected trajectories. This is especially true for offline DA, where no explicit model links
 successive time instants and ensemble members are therefore not necessarily dynamically consistent paths through time
 (Emile-Geay et al., 2025). This issue motivates the emphasis here on temporal coherence, internal uncertainty, structural
 50 uncertainty and proxy composition sensitivity, rather than on visual agreement between ensemble means alone. The CE
 proxy network adds a second layer of complexity because its composition changes strongly through time. Tree rings provide
 dense Northern Hemisphere continental coverage and are essential for large scale temperature reconstructions, but their
 seasonal, ecological and moisture sensitivities can shape annual mean signals. Corals are fewer and more temporally
 clustered, but they sample tropical and subtropical oceans where instrumental hydrological and SST records are short and
 55 where Indo Pacific variability strongly controls regional climate (PAGES 2k Consortium, 2017; Walter et al., 2023). A
 feature that remains stable when these archive components are altered has a different evidential status from one that depends
 on a particular archive configuration.

Here, we use an online Sequential Importance Resampling (SIR) particle filter coupled with Linear Inverse Models (LIMs)
 trained from CMIP5 and CMIP6 simulations to reconstruct annually resolved CE climate fields. The study is designed to
 60 identify which reconstructed features are informative, not only to generate another CE product. We evaluate a reference
 AllProxy reconstruction, then use an 80 member structural ensemble and a 50 member proxy composition ensemble to
 separate posterior spread, methodological sensitivity and observing network sensitivity. Particular attention is given to large



amplitude tropical Pacific and Pacific Walker circulation phases, because these may represent either volcanic responses or internally generated coupled ocean atmosphere states not sampled by the short instrumental window.

65 2 Data and methods

2.1 Online particle filter and LIM framework

We generated annually resolved reconstructions for 1-2000 CE using an online SIR particle filter (Jebri and Khodri, 2023). At each annual assimilation step, the particle filter combines a forecast distribution of the climate state with the available proxy observations for that year. A particle represents one possible realisation of the multivariate climate state. Its likelihood
70 is evaluated from the mismatch between proxy observations and their estimated values derived from the climate state through the relevant PSM. Particle weights are then normalised and the ensemble is resampled, producing a posterior distribution that carries information from previously assimilated years into the following step. This sequential formulation is important for the interpretation of multidecadal variability. Because particles are propagated from one assimilation step to the next, particle histories have a temporal meaning that differs from offline DA products in which each reconstructed year is
75 estimated independently from a prior ensemble. The online formulation does not remove structural or proxy network uncertainty, but it provides a more appropriate framework for asking whether high amplitude Niño3.4, TPDV and PWC phases are temporally coherent features rather than isolated annual states.

The posterior weight of particle k at year i is proportional to the product of the prior particle weight and the likelihood of the observed proxy vector. Under a Gaussian observation error model, the likelihood depends on the difference between the
80 proxy observations and their model equivalent values, scaled by the observation error covariance matrix. The implementation uses 1000 particles at every assimilation year. The error model includes the residual uncertainty of each PSM, combined with a representativeness error corresponding to the EOF truncation. The sequential weighting, error covariance specification and residual resampling scheme are detailed in the particle filter implementation (Appendix A).

Particle propagation is governed by LIMs. Each LIM represents the evolution of a reduced dimensional climate state as a
85 linear stochastic process estimated from a parent climate model simulation. The state vector combines annually averaged near surface air temperature (Tas), precipitation (Precip), evaporation (Evap), sea surface temperature (Tos) and sea level pressure (Slp). Before LIM estimation, the parent model fields are linearly detrended and projected onto an empirical orthogonal function basis. The truncation level is selected to retain the maximum number of modes compatible with a positive semi definite estimate of the stochastic noise covariance matrix. The LIM evolution can be written as:

90

$$x(t + \Delta t) = Bx(t) + \eta(t), \quad \eta(t) \sim N(0, Q) \quad (1)$$

where x denotes the reduced climate state, B is the linear propagator, and η is a stochastic forcing with covariance Q . The matrices B and Q are estimated from long pre existing model simulations following the LIM formulation. Numerical



95 integration is performed using a two step, second order Runge-Kutta scheme (Penland and Matrosova, 1994), with a time
 step of $\Delta t = 1/120$ year corresponding to approximately 3 days. The reconstructed physical fields are obtained by projecting
 the posterior state back into the original spatial basis. The detailed methodology is described in Jebri and Khodri (2023).

2.2 Climate model priors and bias correction

The prior ensemble is represented by ten LIMs. Eight LIMs are trained independently from individual CMIP5 and CMIP6
 100 simulations spanning the pre industrial interval 850-1850 CE. Two additional LIMs are trained from concatenated
 simulations within the CMIP5 and CMIP6 generations, respectively, and are referred to as the multi model LIM priors. The
 use of individual model and pooled model LIMs allows us to assess how model dependent covariance structures and
 dynamical characteristics influence the assimilation outcome. The parent simulations used to train each LIM and the EOF
 truncation levels retained for each prior are listed in the LIM construction summary (Appendix B).
 105 For each prior, reconstructions are carried out with and without a bias correction procedure applied to the LIM based climate
 fields. Bias correction is intended to reduce systematic differences in the spatial anomaly patterns associated with large-scale
 climate modes between the prior model and the observational products used for PSM calibration. Because the correction may
 modify mean states and covariance relationships that are relevant to assimilation, its influence is evaluated explicitly rather
 than treated as a fixed preprocessing step. The decoder based pattern correction, observational target, reference period and
 110 diagnostic example are described in the bias correction method (Appendix C).

2.3 Proxy database and Proxy System Models

The assimilated proxy database merges the records used in the PHYDA compilation (Steiger et al., 2018), including records
 from PAGES 2k Consortium (2017), tree ring width records used in Breitenmoser et al. (2014) and earlier regional
 compilations, with additional records from the Iso2k database (Konecky et al., 2020). Duplicates are removed before
 115 calibration. The final database contains 3386 annual proxy series, with the exact number varying slightly among PSM
 calibrations because only records meeting the calibration requirements are retained. The network is dominated by tree ring
 records, which account for more than 75 % of the retained series and are concentrated over Northern Hemisphere continental
 regions. Coral records are fewer but provide strategically important coverage in the tropical and subtropical oceans. Figure 1
 documents the spatial distribution and temporal availability of the full database.

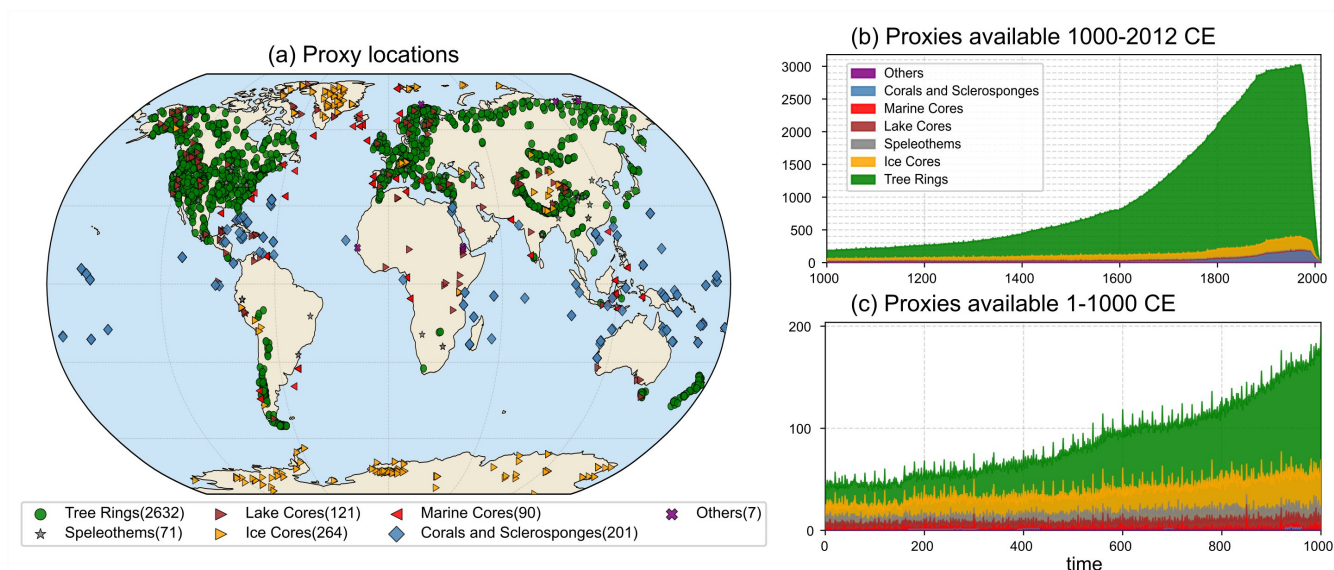


Figure 1: Proxy availability over the Common Era. (a) Spatial distribution of the proxy records available for assimilation. (b) Number of proxy records available between 1000 and 2012 CE. (c) Number of proxy records available between 1 and 1000 CE. The database combines the proxy records compiled by Steiger et al. (2018), including the PAGES 2k database (PAGES 2k Consortium, 2017), with additional records from Iso2k (Konecky et al., 2020).

Proxy observations are related to climate variables through statistical PSMs calibrated over the instrumental period (Tardif et al., 2019). For non-tree ring records, a univariate linear regression against local surface temperature is used. For tree ring records, three candidate models are evaluated: a temperature only regression, a precipitation only regression, and a bivariate temperature and precipitation regression. The retained formulation is selected using the Bayesian Information Criterion in order to balance explanatory power and parsimony. Only proxy series with at least 20 years of overlap with the instrumental calibration data are used. We used two different types of instrumental datasets for calibration. The first uses NOAA 20th Century Reanalysis version 3 that provides spatially complete and dynamically consistent temperature and precipitation fields (Compo et al., 2011; referred here after as NOAA), and the second combines the GISTEMP version 4 surface temperature (Hansen et al., 2010; Lenssen et al., 2019), which provides partial coverage with some persistent gaps at the highest latitudes, and the GPCC version 2022 land precipitation (Becker et al., 2013; Schneider et al., 2022). This second dataset is referred hereafter as GG. For the first calibration dataset, we tested three different calibration periods (1900-2012 referred NOAA1, 1920-2012 referred NOAA2, 1940-2012 referred NOAA3) to assess sensitivity to calibration period length. The spatial distribution and retained numbers of proxy records for the NOAA1, NOAA2, NOAA3 and GG calibration setups are shown in the proxy database comparison (Appendix E, Fig. E). The regression residual variance contributes to the observation error specification used by the particle filter. The tree ring candidate regressions, the univariate PSM used for other archives and the Bayesian Information Criterion used for model selection are given in the PSM formulation (Appendix D). This PSM framework is intentionally simple and consistent across the experiment suite. It does not attempt to resolve all potential non linearities, seasonal sensitivities or non stationarity in proxy climate relationships.



Rather, by varying calibration data sources and calibration periods, the experiments quantify the sensitivity introduced by plausible choices that are routinely made in largely used statistical PSM construction (Steiger et al., 2018; PAGES 2k Consortium, 2019; Valler et al., 2024; Wu et al., 2025).

2.4 Experimental design

We designed two complementary experiment suites (Fig. 2 ; Table 1). The first quantifies structural uncertainty associated with PSM calibration, model prior selection and bias correction. The second assesses proxy composition uncertainty associated with tree ring and coral inclusion. In all experiments, all components not explicitly varied are held fixed, allowing the effect of each choice and its interactions to be isolated.

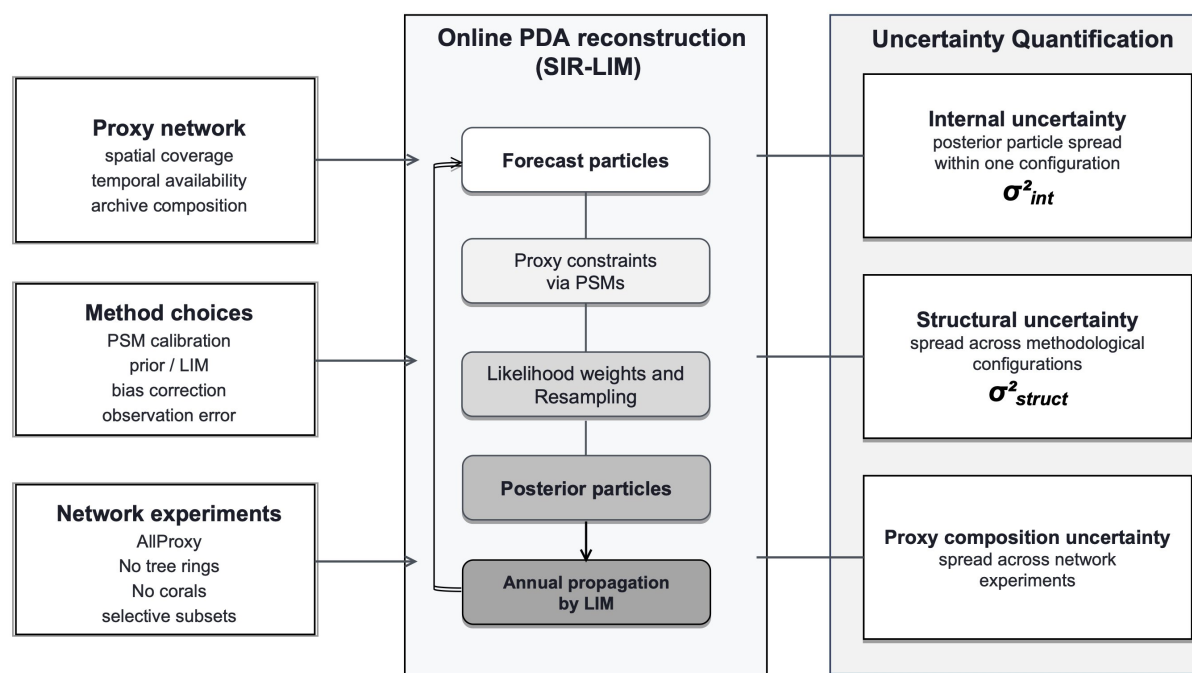


Figure 2: Conceptual framework for separating internal, structural and proxy composition uncertainty in online paleoclimate data assimilation. For each forecast particle, PSMs predict proxy values from the simulated climate state. The mismatch between predicted and observed proxies defines likelihood weights, which are then used in the resampling step to obtain posterior particles.

2.4.1 Structural uncertainty ensemble

The structural ensemble consists of 80 reconstructions obtained from a factorial combination of four PSM calibrations, ten LIM priors and two bias correction options. The PSM calibrations include the three configurations based on NOAA 20th Century Reanalysis version 3 calibration periods 1900-2012, 1920-2012 and 1940-2012, and one configuration based on GISTEMP temperature and GPCC precipitation products. The ten priors comprise eight individual model LIMs and two



multi model LIMs. Each combination is performed with and without bias correction (referred to as $b=1$ and $b=0$ respectively). This design samples uncertainty associated with observational calibration products and periods, model dependent prior dynamics, and systematic prior bias (Table 1).

Experiment component	Configurations	Purpose
PSM calibration	NOAA1 (1900-2012); NOAA2 (1920-2012); NOAA3 (1940-2012); GG (GISTEMP + GPCC)	Assess sensitivity to calibration dataset, spatial coverage and calibration period length
LIM prior	8 single model LIMs 1 LIM CMIP5 1 LIM CMIP6	Assess model dependent covariance and dynamical structure
Bias treatment	$b=1$, with bias correction $b=0$, without bias correction	Assess effect of systematic prior model biases
Total	$4 \times 10 \times 2 = 80$ configurations	Structural uncertainty ensemble

Table 1. Structural uncertainty experiment design. Parent simulations used for the different LIM calibrations, together with the corresponding EOF truncation levels, are listed in the LIM prior summary (Appendix B).

2.4.2 Proxy composition ensemble

Proxy composition experiments are performed with the configuration using NOAA1 PSM calibration and bias corrected priors. For each of the ten priors, five proxy network configurations are assimilated: all proxies, tree rings only, all proxies excluding tree rings, corals only, and all proxies excluding corals. This produces 50 reconstructions (Table 2). The experiments are not intended to determine the intrinsic value of one archive type in isolation. Instead, they diagnose the sensitivity of the reconstruction to the available proxy network and reveal whether selected signals depend strongly on tree rings, corals, or their joint contribution. The spatial contribution of each proxy type is evaluated using a correlation loss diagnostic. For each grid point, variable and prior, the temporal correlation is computed between the AllProxy reconstruction and the reconstruction obtained after excluding the proxy type considered. The contribution is then defined as

$$\text{Contribution} = 1 - \text{Corr}(\text{AllProxy}, \text{NoProxy}) \quad (2)$$

where $\text{Corr}(\text{AllProxy}, \text{NoProxy})$ is calculated over the selected analysis period. Values close to zero indicate that excluding the proxy type has little effect on the reconstructed temporal variability, whereas larger values indicate that the



reconstruction changes more substantially when that proxy type is removed. Stippling identifies grid points where all priors indicate a strong divergence between the AllProxy and exclusion experiments, defined by $r^2 < 0.5$. This metric is interpreted as an archive sensitivity diagnostic and not as a direct measure of reconstruction skill.

Proxy network configuration	Interpretation
All proxies	Reference network for proxy composition experiments
Tree rings only	Information supported by terrestrial ring width or ring density records alone
No tree rings	Sensitivity to excluding the dominant continental archive type
Corals only	Information supported by tropical and subtropical marine records alone
No corals	Sensitivity to excluding marine records that sample key tropical regions

185 **Table 2. Proxy composition experiment design. Ten priors are used for each proxy network configuration, yielding 50 experiments.**

2.5 Evaluation against instrumental and existing reconstruction products

We evaluate a reference reconstruction based on the NOAA1 calibration with bias correction and the all proxy network. The central estimate is the median across the ten prior model reconstructions, while the inter prior range is shown to characterise prior dependence within this reference setup. Index level evaluation considers global mean temperature (GMT), Northern Hemisphere surface temperature (NH, 30°-90°N), North Atlantic sea surface temperature (NASST) estimated as the SST mean over the 75°-7.5°W and 0°-60°N region, the El Niño Southern Oscillation (ENSO) index defined from surface temperature in the Niño3.4 region (McPhaden et al., 2006), the tropical Pacific decadal variability (TPDV) defined as the 10 year low pass filtered TPI index (Henley et al., 2015), and the Pacific Walker circulation (PWC) defined from the trans Pacific equatorial sea level pressure gradient as the difference between the SLP mean anomalies over the central eastern Pacific (160°-80°W, 5°S-5°N) and the western Pacific (80°-160°E, 5°S-5°N).

Spatial reconstruction skill is assessed against instrumental products for Tas, Tos, precipitation, evaporation and SLP. These include Berkeley Earth surface temperature for Tas (Rohde and Hausfather, 2020), ERSSTv5 and HadISST for Tos (Huang et al., 2017; Rayner et al., 2003), GPCC and GPCP for precipitation (Schneider et al., 2022; Adler et al., 2003), NOAA20CRv3 for evaporation (Slivinski et al., 2019; Compo et al., 2011), and HadSLP2 and NOAA20CRv3 for SLP (Allan and Ansell, 2006; Slivinski et al., 2019; Compo et al., 2011), as appropriate for each variable and overlap period. Correlation and root mean square error (RMSE) are calculated from anomalies relative to a 1960-1990 baseline. Correlation estimates the temporal phase alignment between two signals, independently of their amplitude while the RMSE quantifies the overall magnitude of their mismatch. Detrended correlations are additionally evaluated to distinguish agreement associated with common long term forcing from skill in shorter timescale variability. Because some products may overlap with calibration data, these diagnostics are interpreted as evaluation against observational references.

For CE context, the reconstruction is compared with PHYDA (Steiger et al., 2018), the PAGES 2k multi method reconstructions (PAGES 2k Consortium, 2019), Mode-RA (Valler et al., 2024), LMRv3 (Perkins and Hakim, 2021) and



Falster et al., 2023 over their common intervals. Inter-product comparison is used to identify consistent behaviour and to highlight intervals or indices for which plausible products disagree. It is not used to establish a unique reference truth before the instrumental period.

2.6 Uncertainty decomposition and attribution

For a given reconstruction configuration c , year t and scalar index or grid point s , let x denote the posterior particle distribution and $\mu_c(t,s)$ its mean. Internal uncertainty is the posterior particle variance conditional on that configuration, averaged across configurations:

$$\sigma_{\text{int}}^2(t,s) = E_c [\text{Var}(x(t,s) | c)] \quad (3)$$

Structural uncertainty is the variance of configuration means across the 80 member factorial ensemble:

$$\sigma_{\text{struct}}^2(t,s) = \text{Var}_c [\mu_c(t,s)] \quad (4)$$

The total variance is represented as the sum of these terms:

$$\sigma_{\text{total}}^2(t,s) = \sigma_{\text{int}}^2(t,s) + \sigma_{\text{struct}}^2(t,s) \quad (5)$$

This decomposition distinguishes uncertainty that remains after methodological choices have been fixed from uncertainty induced by varying those choices. For the proxy network experiments, we calculate an analogous proxy composition uncertainty from the variance among proxy network configuration means (Eq. (6)), retaining the designation proxy composition uncertainty rather than structural uncertainty.

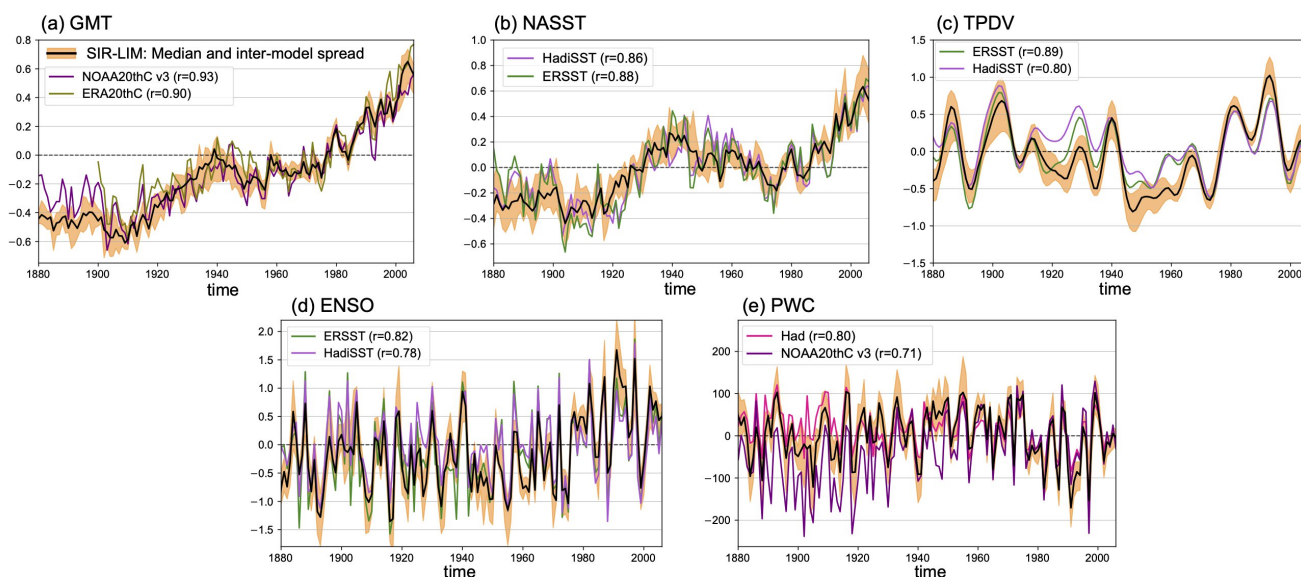
$$\sigma_{\text{proxy}}^2(t,s) = \text{Var}_p [\mu_p(t,s)] \quad (6)$$

We use analysis of variance (ANOVA) to partition structural variance among the main effects of PSM calibration, prior model and bias correction, together with their interactions. The analysis is performed for the full band signal and for frequency bands representing interannual (3-8 years), decadal (8-30 years), multidecadal (30-100 years), and secular (>100 years) variability. The ANOVA fractions are interpreted as relative contributions to configuration-to-configuration variance under the sampled experimental design. They do not represent universal uncertainty fractions beyond the range of configurations evaluated here.

3 Results

3.1 Instrumental period evaluation

We first assess the reference reconstruction during the instrumental period, when proxy coverage is densest and observational products provide the most direct benchmark (Fig. 3). Large scale temperature indices are reproduced with the strongest agreement. The reconstructed GMT and NASST capture the broad twentieth century warming evolution, including the mid century slowdown, and their low frequency behaviour is consistent with the corresponding instrumental products, with correlations above 0.9 and 0.86, respectively. The reconstructed TPDV index also reproduces much of the observed decadal evolution, with correlations of 0.8 and 0.89 with HadISST and ERSST, respectively.



240 **Figure 3: Instrumental period evaluation of reconstructed climate indices. (a) Global mean surface temperature (GMT , °C), (b)**
 245 **North Atlantic sea surface temperature (NASST , °C), (c) tropical Pacific decadal variability (TPDV , °C), defined as the 10 year**
low pass filtered Tripole Index of Henley et al. (2015), (d) ENSO variability, represented by the Niño3.4 index (°C), and (e) Pacific
Walker circulation (PWC) variability, estimated from the trans-Pacific equatorial sea level pressure gradient (Pa). Results are
shown for the NOAA1(b=1) reference experiment with bias correction compared with several instrumental products. Correlations
between the inter prior median reconstruction and each instrumental product over 1900-2004 are indicated in the legend. Yellow
shading represents the spread across reconstructions derived from the different LIM priors, including both single model priors
and multi model priors, LIM CMIP5 and LIM CMIP6.

The tropical Pacific and circulation indices require a more qualified interpretation. Niño3.4 variability shows agreement in the timing of several observed events, but the reconstructed amplitude is less tightly constrained and remains sensitive to the configuration. The reconstructed PWC is consistent with observed products from approximately 1920 onward. Before that date, observational uncertainty itself is large, and the reconstruction lies within the spread between available pressure products rather than providing a definitive estimate of the true circulation state. These results indicate that the online system can retain dynamically coherent tropical signals, while also confirming that their uncertainty is larger than for broad scale temperature indices.

255 Spatial evaluation supports these conclusions (Fig. 4). Non detrended correlations show the strongest skill for Tas and Tos, especially in regions with dense proxy coverage and in dynamically connected oceanic regions (Fig. 4a and b). Skill is generally higher in the Northern Hemisphere mid latitudes and in parts of the tropics than in the sparsely constrained Southern Hemisphere. Precipitation and evaporation show more heterogeneous correlations (Fig. 4c-d), reflecting both their stronger local variability and the limitations of annual proxy constraints. Nevertheless, useful correlations in SLP (Fig. 4e)

260 demonstrate that the assimilation system can propagate information across dynamically related variables and regions that are not directly observed by proxies. After detrending, correlations remain substantial in selected regions, particularly in the Pacific and North Atlantic (Fig. 4f-j), showing that the agreement is not solely produced by the shared twentieth century



warming trend. The archive specific contribution to spatial skill during the instrumental period is further assessed using exclusion experiments for tree rings and corals across Tas, Tos, precipitation and SLP (Appendix F, Fig. F). This comparison helps identify where skill is primarily supported by the dominant terrestrial archive network and where sparse marine records make a detectable contribution. At the same time, the spatial evaluation highlights that skill is variable specific and geographically uneven. Tree ring records contribute most clearly to surface air temperature and, to a lesser extent, sea level pressure over extratropical continental regions, whereas coral records make a more detectable contribution to tropical sea surface temperature, rainfall and tropical Pacific related atmospheric fields. Consequently, the uncertainty analyses of the following sections focus on large scale indices and explicitly consider spatial patterns of proxy influence rather than assuming uniform reconstruction quality over the global domain.

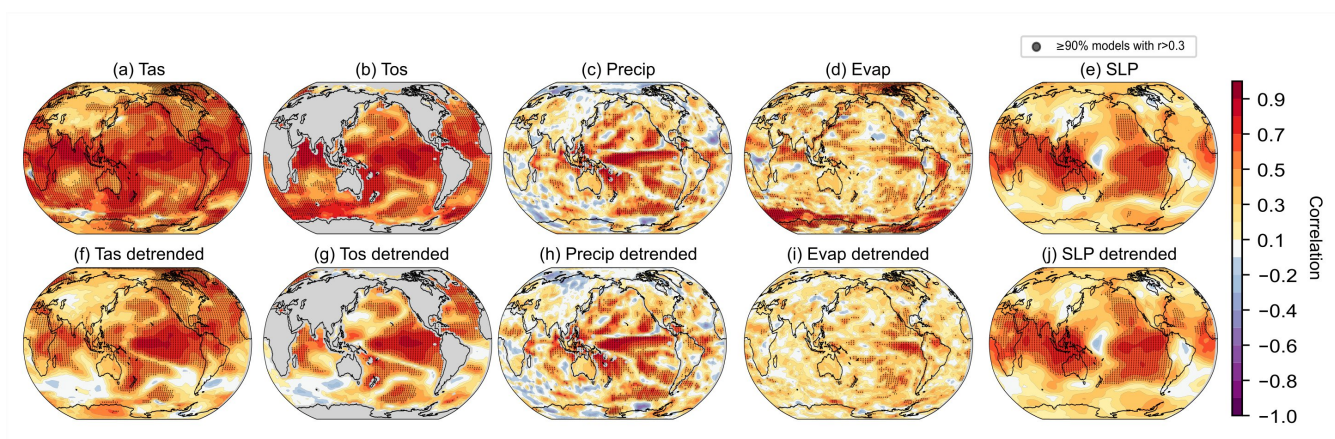


Figure 4: Spatial correlation with instrumental products. Correlation scores between the median NOAA1 (b=1) reconstruction and Berkeley Earth for Tas, ERSSTv5 for Tos, GPCC for land precipitation, GPCP for ocean precipitation, NOAA20CRv3 for evaporation and HadSLP2 for SLP. Correlations are calculated over 1900–2000, except for ocean precipitation, which is evaluated over 1979–2000. The first row shows correlations before detrending and the second row after detrending. Dots indicate grid points where at least 90 % of the LIM prior reconstructions have a correlation greater than 0.3.

3.2 Common Era reconstruction and comparison with existing products

The reference reconstruction provides a coherent view of CE variability while also showing where product comparison alone is insufficient (Fig. 5). GMT displays the expected transition from pre industrial variability to industrial era warming and lies within the spread of available reconstructions over the last millennium. This agreement is important but not surprising, given the dense recent proxy network and the large forced signal. The more informative comparison concerns departures among products before the instrumental period, because they indicate which reconstructed features require explicit uncertainty attribution rather than simple visual comparison (Appendix G, Fig. G).

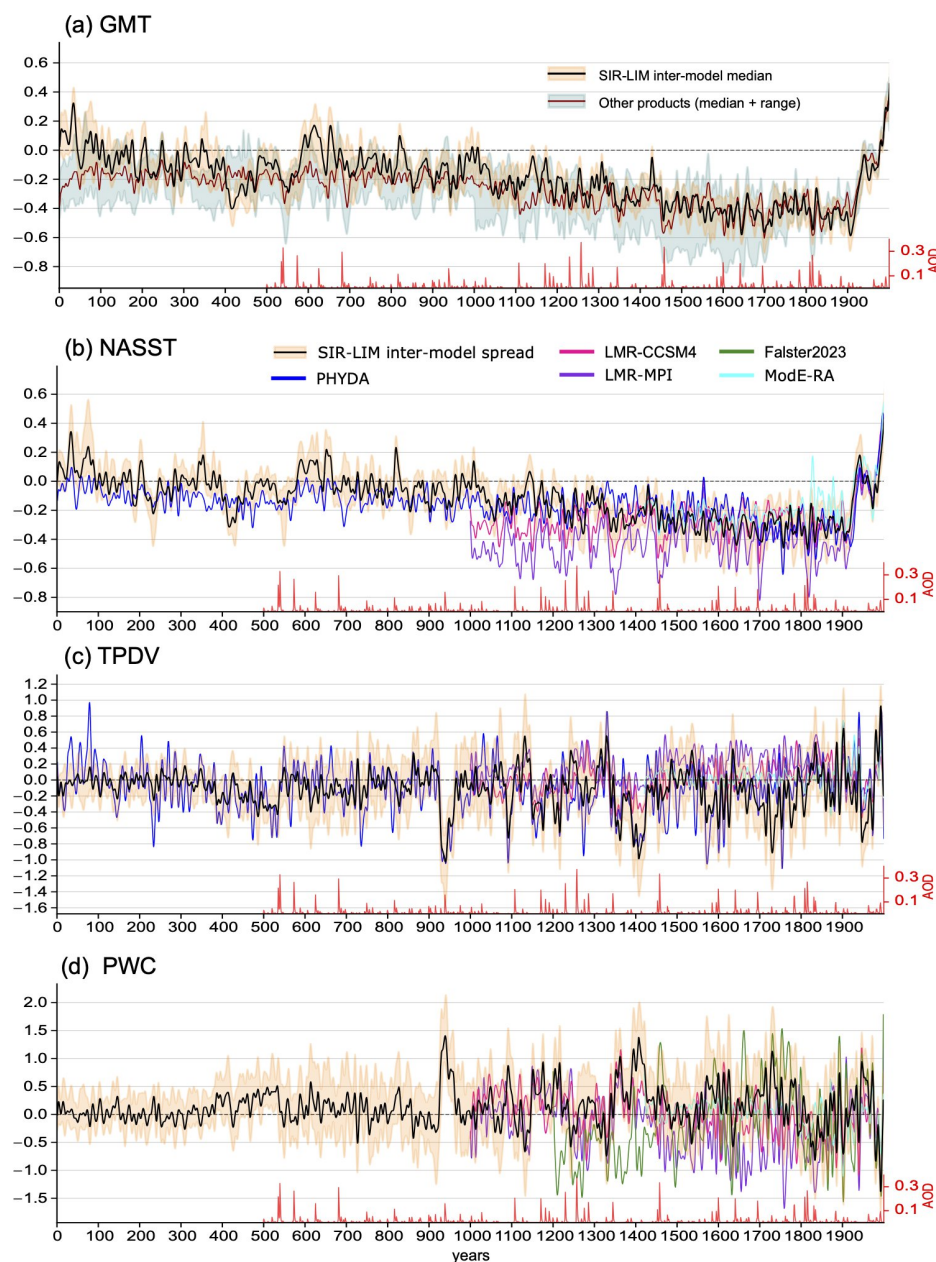


Figure 5: Reconstructed climate indices over the last two millennia. Indices from the NOAA1 reference experiment with bias correction for (a) global mean surface temperature (GMT, °C), (b) North Atlantic sea surface temperature (NASST, °C), (c) tropical Pacific decadal variability (TPDV, °C), and (d) the standardised Pacific Walker circulation (PWC) index. Anomalies are expressed relative to the 1960-1990 reference period. Coloured lines show comparison with existing paleoclimate reconstruction products. The orange shading represents the spread across the different LIM priors, including both single model priors and multi model priors, LIM CMIP5 and LIM CMIP6. A 10 year low pass filter was applied to all time series.



The reconstruction of NASST shows a gradual cooling from the early second millennium culminating toward the LIA, followed by modern warming (Fig. 5b). This evolution is broadly comparable with PHYDA at some timescales, but the magnitude and timing of secular North Atlantic changes differ among reconstruction products. PHYDA shows a more abrupt late LIA change, whereas LMR products display less pronounced secular MCA-LIA contrasts in the corresponding index. These differences illustrate why NASST is an important target for structural uncertainty analysis: the broad low frequency evolution is meaningful, but product disagreement cannot be interpreted solely as random uncertainty when plausible methodological choices can shift the amplitude of secular variability.

TPDV is more coherent across products than NASST at decadal to multidecadal timescales (Fig. 5c). The present reconstruction displays marked phases around approximately 500, 950, 1150 and 1400 CE that are also visible in PHYDA, with a statistically meaningful correlation over the common two millennia interval ($r = 0.63$). However, the amplitude of our reconstruction changes more strongly with proxy coverage, whereas PHYDA retains more uniform variability through proxy sparse periods. This difference is not a reason to privilege one product, but a diagnostic of how prior dynamics, proxy treatment and uncertainty regularisation can control tropical Pacific amplitude.

The PWC reconstruction is less consistent across products, especially before 1500 CE (Fig. 5d). It shows partial phase agreement with LMR over the last millennium and remains dynamically anti correlated with TPDV in the present system ($r = -0.96$), but differs more strongly from the Falster et al. (2023) reconstruction. The key question is therefore when a reconstructed PWC anomaly is supported by the assimilation rather than imposed by product specific assumptions. Sections 3.3-3.5 address this by asking whether large PWC phases coincide with temporal coherence, reduced internal uncertainty, reduced structural uncertainty and persistence across proxy composition experiments. This strategy is consistent with recent recommendations that temporal comparisons among PDA products should not rely only on ensemble mean proximity or the apparent overlap of uncertainty plumes (Emile-Geay et al., 2025). In the following sections, product agreement is therefore used only as a contextual indicator; the stronger evidence for robustness comes from the joint behaviour of temporal coherence, internal uncertainty, structural uncertainty and proxy composition experiments.



3.3 Temporal coherence across methodological configurations

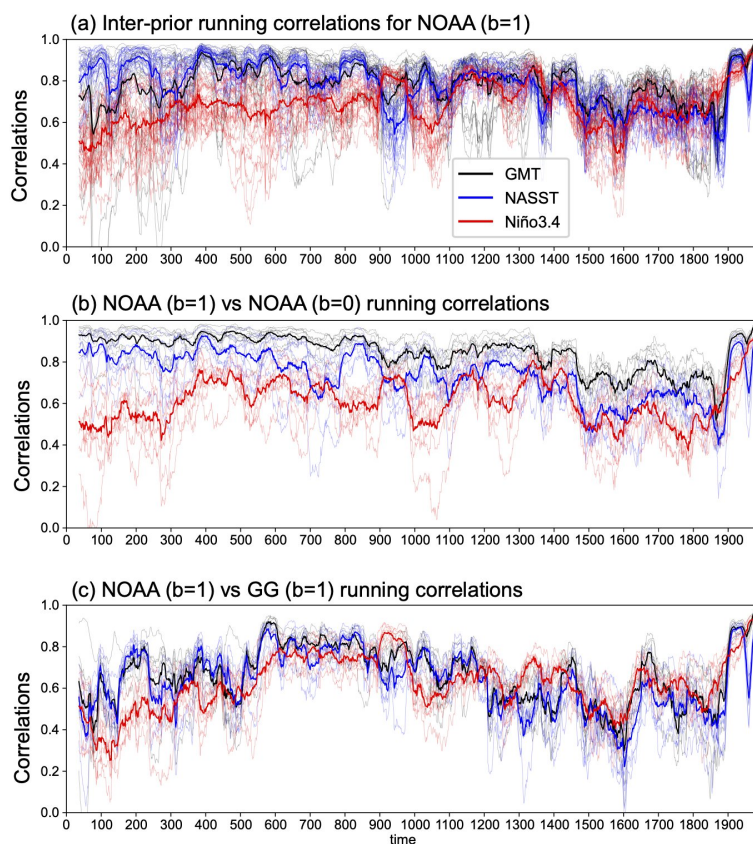


Figure 6: Temporal consistency across methodological configurations. Seventy year running correlations for global mean surface temperature (GMT, black), North Atlantic sea surface temperature (NASST, blue) and the Niño3.4 index (red). Thin lines show individual pairwise correlations, and bold lines show the corresponding means. (a) Inter model consistency, computed from all possible pairs of LIM priors within the NOAA1(b=1) experiment, measuring the robustness of the reconstruction across different priors. (b) Effect of bias correction, computed between NOAA1 (b=1) and NOAA1 (b=0) for each prior, quantifying the impact of the bias correction. (c) Effect of the PSM calibration, computed between NOAA1 (b=1) and GG (b=1) for each prior, quantifying reconstruction sensitivity to changes in the PSM calibration.

We next use 70 year rolling correlations to distinguish temporal coherence from skill against observations (Fig. 6). GMT and NASST remain relatively coherent across LIM priors, bias correction choices and PSM calibrations, indicating that their broad temporal phasing is not controlled by one configuration alone. Niño3.4 is less coherent and more sensitive to both prior choice and bias correction, especially during proxy sparse intervals. This contrast is physically consistent with the fact that tropical Pacific variability depends on coupled ocean-atmosphere dynamics and teleconnections that are only indirectly sampled by much of the proxy network. The coherence diagnostics therefore provide a filter for interpreting the variance decomposition. Periods of high coherence do not necessarily imply small amplitude uncertainty, and periods of low average coherence do not make every reconstructed feature unusable. Instead, the diagnostic identifies where phase stability,

amplitude uncertainty and proxy constraint must be examined together. This is particularly important for detecting high amplitude tropical Pacific and PWC phases that may be robust over limited intervals even though the full time series remains difficult to reconstruct continuously.

3.4 Internal and structural uncertainty

The uncertainty decomposition separates posterior particle spread from disagreement among methodological configurations (Figs. 7 and 8). For GMT and NASST, both components are similar before 1200 CE, when proxy coverage is sparse and posterior constraint is weaker (Fig. 7a, b, d, e). After 1200 CE, internal uncertainty decreases, but structural uncertainty remains substantial. Combined with the temporal coherence results, this means that GMT and NASST contain reliable information on the timing and ordering of major low frequency features, whereas their amplitude remains partly method dependent. The frequency decomposition identifies why this amplitude limitation persists. Structural uncertainty in GMT and NASST is dominated by secular variability and mainly controlled by PSM calibration, with secondary contributions from prior choice and PSM prior interactions (Appendix I, Fig. I). This result echoes the broader finding that Common Era reconstructions can be more consistent at multidecadal timescales than at multicentennial amplitude (PAGES 2k Consortium, 2019). In our ensemble, the 400-600 CE cold episode, the 600-700 CE warm optimum and late LIA cooling are therefore interpretable primarily as coherent low frequency features whose exact magnitude remains calibration dependent.

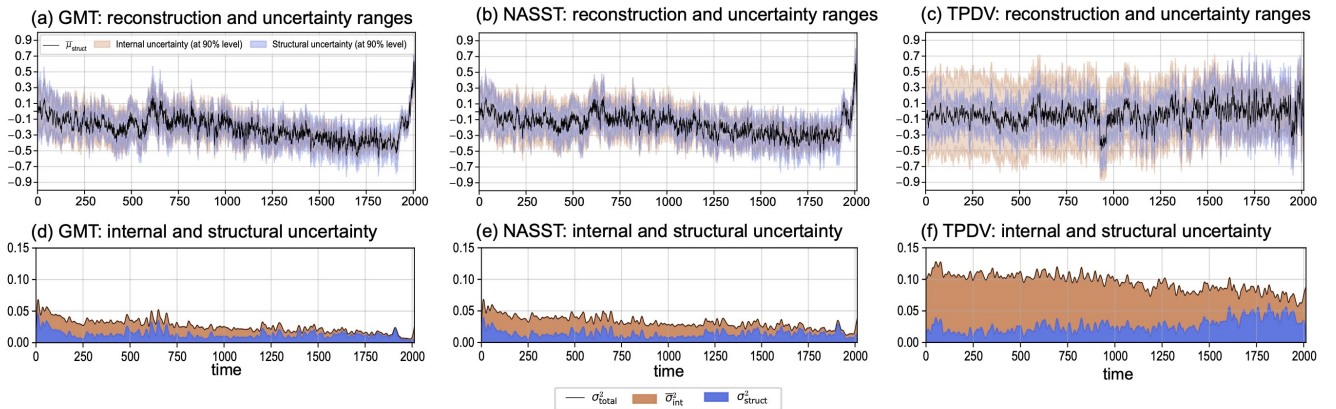


Figure 7: Index level uncertainty partition for GMT, NASST and TPDV. Uncertainty partition for reconstructed global mean surface temperature (GMT, °C; first column), North Atlantic sea surface temperature (NASST, °C; second column) and tropical Pacific decadal variability (TPDV, °C; third column), using the AllProxy reconstruction ensemble. (a-c) Reconstructions and associated uncertainty ranges, computed across the 80 tested configurations, which differ in PSM calibration, LIM prior and bias correction. Black lines show the inter configuration ensemble mean; orange shading shows the mean internal uncertainty range at the 90 % confidence level, averaged across configurations; blue shading shows the structural uncertainty range at the 90 % confidence level across configurations. (d-f) Total uncertainty, represented as the sum of the mean internal variance, σ_{int}^2 , and the structural variance, σ_{struct}^2 , together with their respective contributions to total reconstruction uncertainty.



TPDV shows a different uncertainty balance (Fig. 7c, f). Its total uncertainty remains large, with a greater contribution from internal uncertainty than for GMT and NASST, even though the reconstructed decadal evolution agrees relatively well with PHYDA. Product agreement therefore supports the timing of some decadal features but does not imply that the instantaneous tropical Pacific state is tightly constrained. The structural component is also more prior dependent and mainly expressed at decadal to multidecadal timescales (Appendix I, Fig. I), consistent with the role of prior covariance in projecting sparse proxy information into the tropical Pacific.

The strongest tropical Pacific result emerges for Niño3.4 and PWC (Fig. 8). Although internal uncertainty dominates over much of the CE, the largest reconstructed phases are not always the least reliable. Around approximately 950, 1200 and 1400 CE, high amplitude Niño3.4 and PWC excursions coincide with marked reductions in posterior spread and reduced structural uncertainty (Fig. 8a-c, b-d; Appendix J, Fig. J). These intervals therefore stand out as constrained dynamical states rather than as large excursions embedded in uniformly high uncertainty.

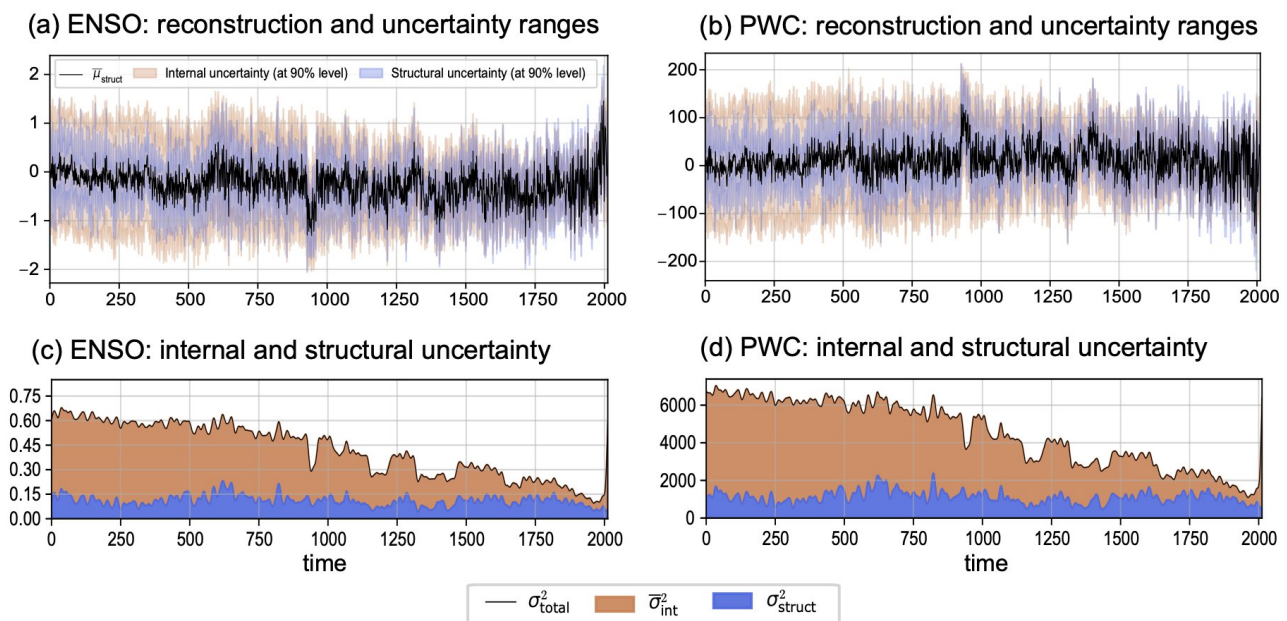


Figure 8: Index level uncertainty partition for Niño3.4 and PWC. Same as Fig. 7, but for the Niño3.4 index (°C, left column) and the Pacific Walker circulation (PWC) index (Pa, right column). The figure shows the reconstruction uncertainty ranges and the partition between mean internal variance and structural variance.

This behaviour changes the interpretation of tropical Pacific variability in the reconstruction. The instrumental period provides no direct analogue for these high amplitude, reduced uncertainty phases, and the structural decomposition indicates that their robustness is concentrated at interannual to multidecadal timescales rather than in a smooth secular trend (Appendix J, Fig. J). These phases could reflect volcanic forcing, internally generated coupled ocean-atmosphere variability, or a state dependent interaction between forcing and internal variability. The present experiments do not attribute the events, but they identify them as priority targets for mechanism oriented analysis.



Overall, the structural analysis separates two types of information. GMT and NASST provide coherent low frequency features whose amplitudes are limited mainly by PSM calibration. TPDV, Niño3.4 and PWC provide event scale information when large reconstructed anomalies coincide with reduced internal and structural uncertainty. This distinction is tested further with the proxy composition experiments.

3.5 Proxy composition uncertainty and archive specific contributions

We now assess proxy composition uncertainty, defined as the sensitivity of the reconstruction to the inclusion or exclusion of tree rings and corals. Unlike the structural ensemble, these experiments modify the observing network itself. They therefore capture uncertainty arising from archive availability, seasonal sensitivity and spatial sampling, rather than from changes in the assimilation framework.

3.5.1 GMT, NASST and TPDV

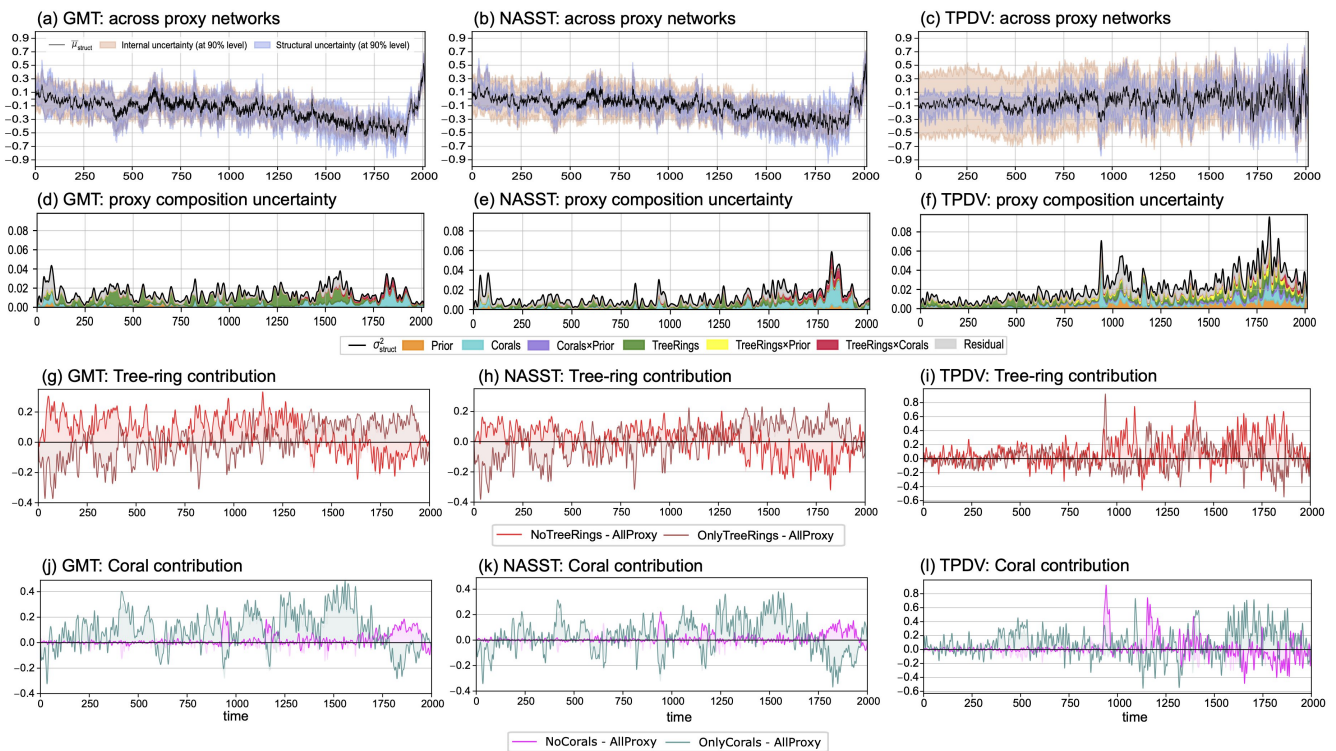


Figure 9: Proxy composition uncertainty and archive specific contributions for global mean surface temperature (GMT, °C ; first column), North Atlantic sea surface temperature (NASST, °C ; second column) and tropical Pacific decadal variability (TPDV, °C ; third column). (a-c) Reconstructed indices from the sensitivity experiments described in Table 2. Black lines show the ensemble mean averaged across the 50 proxy composition configurations, which differ in assimilated proxy type and LIM prior. Orange shading shows the mean internal uncertainty range at the 90 % confidence level, and blue shading shows proxy composition uncertainty arising from proxy network composition, expressed as the 5th-95th percentile range. (d-f) ANOVA decomposition of the relative contributions of prior choice and the inclusion or exclusion of coral and tree ring proxies. (g-i) Tree



395 **ring contributions and (j-l) coral contributions ($^{\circ}\text{C}$), estimated as the difference between each selective proxy experiment and the AllProxy reconstruction.**

For GMT, NASST and TPDV, internal uncertainty and proxy composition uncertainty evolve differently through time (Fig. 9a-c). GMT and NASST illustrate a key point: increasing proxy availability can narrow posterior spread while making archive type differences more visible. The result is not a simple monotonic gain in certainty, but a shift in what limits interpretation. For TPDV, both components are larger and more variable, especially after 1500 CE, confirming that tropical Pacific variability is more sensitive to the changing composition of the observing network (Fig. 9c).

The ANOVA decomposition shows that proxy composition effects are episodic rather than controlled by one persistent archive term (Fig. 9d-f). For GMT and NASST, tree ring or coral contributions become important during selected intervals, whereas for TPDV the variance is larger and more event focused over the last millennium. Proxy composition uncertainty therefore acts as an archive consistency test: it identifies when the AllProxy reconstruction is supported by a stable combination of archives and when the result depends more strongly on which proxy class carries the information.

The archive specific contribution panels provide a more direct view of how tree rings and corals affect the reconstructed indices (Fig. 9g-l). For GMT and NASST, tree ring exclusion and tree ring only experiments often produce departures of opposite sign relative to the AllProxy reconstruction (Fig. 9g, h). This approximate symmetry indicates that tree rings actively shape low frequency temperature variability rather than simply reducing posterior spread. Because tree rings represent a large fraction of the assimilated network, their selective sensitivity to growing season temperature or moisture can strongly affect low frequency GMT and NASST variability. For TPDV, tree ring effects are more episodic and less clearly symmetric, suggesting a weaker or more indirect influence on tropical Pacific variability (Fig. 9i).

Coral contributions show a different behaviour. For GMT and NASST, coral only experiments can differ substantially from the AllProxy reconstruction, whereas excluding corals generally produces smaller but detectable changes (Fig. 9j, k). This asymmetry suggests that coral records contain information that is not fully reproduced by the broader land dominated network, or that their signal changes when combined with other proxy types. Before approximately 1700 CE, the coral only response is often not matched by an equivalent no coral response, consistent with sparse and geographically limited coral coverage. After approximately 1700 CE, especially during the late LIA and the transition to the industrial period, the response becomes more symmetric: coral only experiments show colder anomalies, while no coral experiments shift warmer. This indicates that available corals pull the assimilated solution toward a colder late LIA state than the land dominated network alone. The mechanism may involve land-ocean differences, archive specific seasonal sensitivity, a tropical bias of coral coverage, or changes in coral availability through time, but these experiments cannot separate these effects.

For TPDV, coral contributions are more directly expressed, with stronger deviations during selected intervals after approximately 1500 CE and around earlier events where tropical marine coverage is influential (Fig. 9l). These departures indicate that corals can affect tropical Pacific variability, but not in a way that is uniformly dominant through time. Together, the tree ring and coral sensitivity experiments suggest that corals provide a geographically sparse but dynamically important constraint on TPDV, whereas tree rings primarily constrain continental and extratropical temperature signals.



3.5.2 Tropical Pacific and circulation indices

The proxy composition experiments confirm that Niño3.4 and PWC are less tightly constrained than broader temperature indices (Fig. 10). For both indices, internal uncertainty remains larger than proxy composition uncertainty over most of the CE, reflecting the difficulty of reconstructing annual tropical Pacific variability and atmospheric circulation from sparse tropical proxy information (Fig. 10a, b). Proxy composition uncertainty nevertheless increases after approximately 1500 CE, when the denser network makes archive type differences more influential.

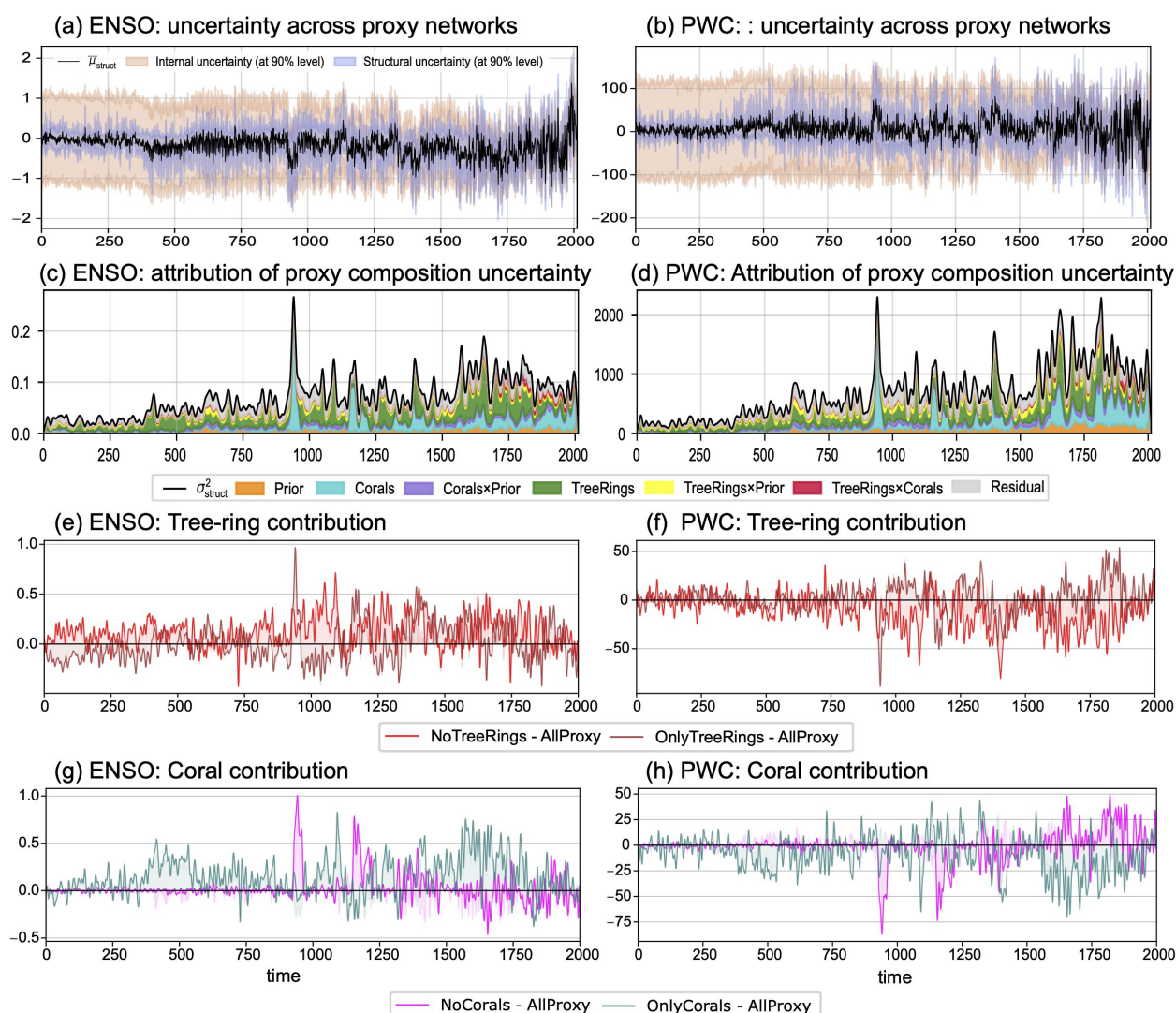


Figure 10: Same as Fig. 9, but for the Niño3.4 index (°C, first column) and the Pacific Walker circulation (PWC) index (Pa, second column). Proxy composition uncertainty and archive specific contributions are shown for Niño3.4 and PWC across the 50 proxy



composition configurations, which differ in assimilated proxy type and LIM prior. Panels (a, c, e, g) show Niño3.4, and panels (b, d, f, h) show PWC. The figure shows the reconstructed indices, the partition between mean internal uncertainty and proxy composition uncertainty, the ANOVA decomposition of the relative contributions of prior choice and the inclusion or exclusion of coral and tree ring proxies, and the archive specific contributions ($^{\circ}\text{C}$) estimated as the difference between each selective proxy experiment and the AllProxy reconstruction.

The ANOVA and archive specific panels show that these proxy composition effects are episodic (Fig. 10c-h). Niño3.4 shows limited proxy composition variance before 1500 CE except during selected events, whereas PWC variance is larger and more intermittent, with peaks during the last millennium and after 1500 CE (Fig. 10c, d). Tree ring and coral departures from AllProxy are also concentrated in selected multidecadal intervals, especially around 950, 1200 and 1400 CE (Fig. 10e-h). This confirms that archive influence is period dependent rather than persistent through time.

The comparison with structural and internal uncertainty gives these archive experiments their main interpretive value. The 950, 1200 and 1400 CE Niño3.4 and PWC phases persist across proxy composition experiments and coincide with reduced uncertainty in Sect. 3.4. They therefore remain meaningful even though their exact amplitudes depend on prior choice and archive distribution. Other intervals show archive sensitivity without the same reduction in internal or structural uncertainty and should be interpreted more cautiously.

3.5.3 Spatial proxy contributions

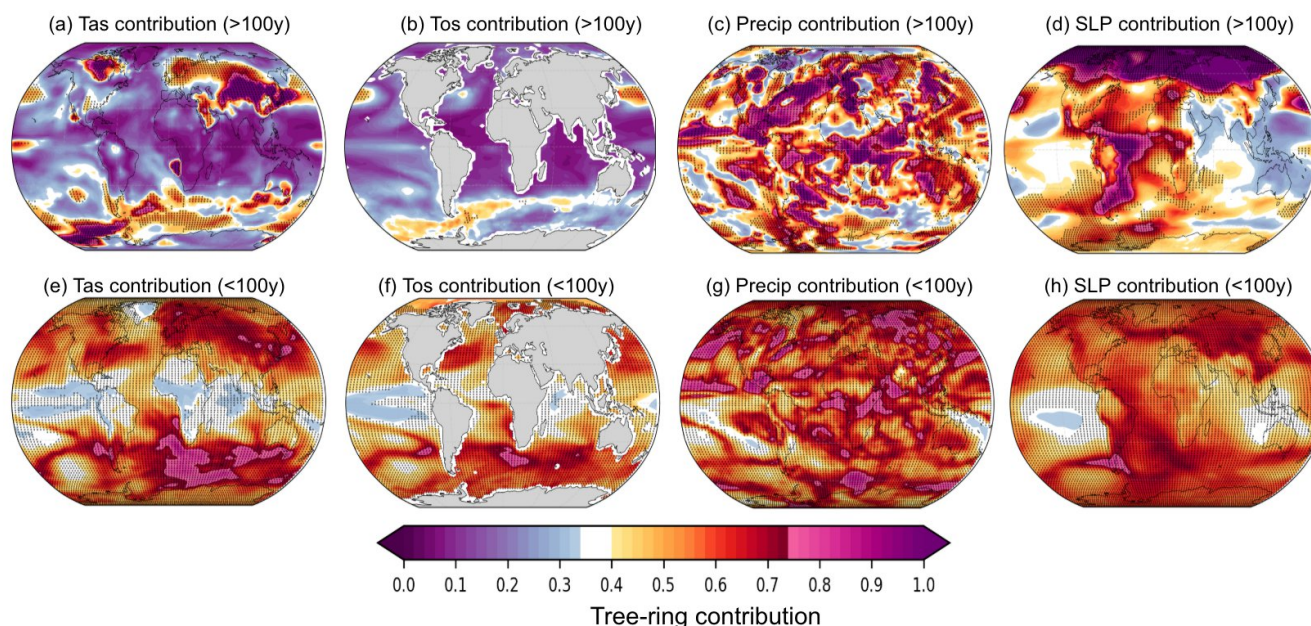


Figure 11: Spatial contribution of tree ring proxies by frequency band over 1000-2000 CE to reconstructed surface air temperature (Tas; a, e), sea surface temperature (Tos; b, f), precipitation (Pr; c, g) and sea level pressure (SLP; d, h). Contributions are decomposed into secular variability, >100 years, in the first row, and interannual to multidecadal variability, <100 years, in the second row. Contributions are estimated from the correlation loss between the AllProxy reconstruction and the corresponding tree ring exclusion experiment. Values close to zero indicate negligible influence, whereas larger values indicate a stronger contribution of tree ring proxies to the reconstructed signal in the corresponding frequency band. Stippling identifies grid points where all priors indicate a strong divergence between the AllProxy and exclusion experiments, defined by $r^2 < 0.5$



The spatial diagnostics show how archive composition affects the reconstructed state at different frequencies (Figs. 11 and 12). Tree ring sensitivity is strongest over Northern Hemisphere continents, where the network is densest, but it also appears in extratropical oceans, precipitation and SLP (Fig. 11). This indicates that the dominant archive class helps determine which large scale covariance patterns are activated in the AllProxy reconstruction, especially at secular and interannual to multidecadal timescales.

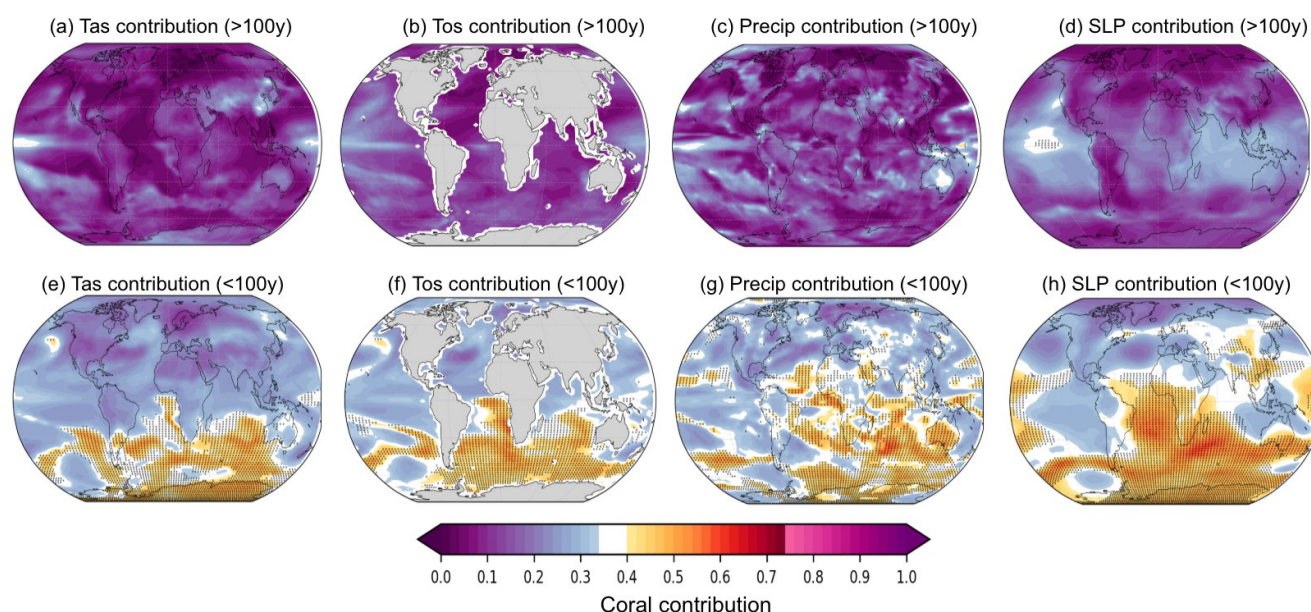


Figure 12: Spatial contribution of coral proxies by frequency band over 1000-2000 CE to reconstructed surface air temperature (Tas; a, e), sea surface temperature (Tos; b, f), precipitation (Pr; c, g) and sea level pressure (SLP; d, h). Contributions are decomposed into secular variability, >100 years, in the first row, and interannual to multidecadal variability, <100 years, in the second row. As in Fig. 11, contributions are estimated from the correlation loss between the AllProxy reconstruction and the corresponding coral exclusion experiment. Values close to zero indicate negligible influence, whereas larger values indicate a stronger contribution of coral proxies to the reconstructed signal in the corresponding frequency band. Stippling identifies grid points where all priors indicate a strong divergence between the AllProxy and exclusion experiments, defined by $r^2 < 0.5$

Coral contributions have a different spatial and frequency structure (Fig. 12). Their influence is lower in amplitude and more diffuse at secular timescales, but becomes more focused over tropical, subtropical and Southern Ocean regions at interannual to multidecadal timescales. This behaviour is consistent with the role of corals as sparse but strategically located archives of tropical ocean temperature and hydrology (Walter et al., 2023). Their contribution is therefore most relevant for interpreting NASST, TPDV, Niño3.4 and PWC features rather than for stabilizing the global mean alone. The spatial patterns further show that coral influence is primarily expressed over ocean regions. This may contribute to the cooler GMT anomalies obtained in coral only experiments in the previous section, because a network dominated by oceanic constraints can weight land sea contrasts differently from the full proxy network.



485 Together, the tree ring and coral contribution maps show that proxy composition can affect both large scale indices and regional fields. Tree rings mainly stabilize continental and extratropical temperature information, whereas corals provide sparse but dynamically important marine constraints. For the AllProxy reconstruction, robust features are therefore those that persist not only across methodological perturbations, but also across changes in which parts of the observing network are allowed to constrain the assimilated state.

490 **4 Discussion**

The product comparison, coherence diagnostics and uncertainty experiments lead to a single interpretation: the reconstruction is informative, but not uniformly so. GMT behaves as expected for a large scale thermodynamic index. It agrees with instrumental products and existing reconstructions during the recent period, and its pre industrial evolution remains broadly comparable with previous CE syntheses. NASST retains a coherent low frequency structure but shows
 495 larger amplitude differences among products. TPDV and PWC differ more strongly from one product to another, not because they are unusable, but because their reconstruction depends on sparse tropical constraints and on how each method represents coupled ocean-atmosphere covariance.

This hierarchy is consistent with the wider literature but sharpens it. PAGES 2k Consortium (2019) showed that global multidecadal temperature variability is relatively coherent across reconstruction methods and simulations, whereas Laepple
 500 et al. (2023) emphasized that regional variability at supradecadal timescales is more likely to be underestimated or misrepresented by models. Our results sit between these two views. GMT confirms the relative robustness of global scale temperature information, while NASST, TPDV and PWC show why regional oceanic and circulation variability cannot be inferred from product agreement alone. It also sharpens how comparisons with other PDA products should be used. Emile-Geay et al. (2025) show that offline DA ensembles require particular care because ensemble members are not necessarily
 505 time connected trajectories, and because visual comparison of plumes can either overstate or understate compatibility. Our interpretation therefore does not rest on the visual overlap of our reconstruction with PHYDA, LMR or PAGES 2k products. Instead, we treat external products as contextual benchmarks and use our own multi configuration diagnostics to decide when a feature has temporal and uncertainty support.

The key advance comes from putting temporal coherence (Sect. 3.3) and uncertainty quantification (Sect. 3.4) together.
 510 Temporal coherence across configurations shows whether a feature has stable phasing; structural uncertainty shows whether its amplitude survives changes in calibration, prior and bias correction. For GMT and NASST, these diagnostics separate timing from magnitude. The 600-700 CE warm optimum, the preceding 400-600 CE cold episode and the late LIA cooling are coherent low frequency features, but their amplitude remains limited mainly by PSM calibration. This explains why large scale reconstructions can agree on broad evolution while disagreeing on the strength of pre industrial anomalies.

515 For the tropical Pacific and PWC, the important result is not simply that uncertainty is larger. The reconstruction identifies intervals when large amplitude Niño3.4, TPDV and PWC phases coincide with reductions in both internal and structural uncertainty. The episodes around approximately 950, 1200 and 1400 CE therefore have a different status from the



surrounding background variability. They are intervals during which the proxy information, prior covariance and methodological ensemble jointly constrain the reconstructed dynamical state.

520 This is scientifically important because these phases have no direct analogue in the short instrumental record. They may reflect the climate system response to volcanic forcing, but they may also represent internally generated coupled ocean–atmosphere states that have not been observed since the nineteenth century. The distinction matters. Volcanic forcing has been linked to multidecadal pre industrial temperature variability, to tropical Pacific changes and to substantial perturbations of tropical hydroclimate (PAGES 2k Consortium, 2019; Khodri et al., 2017; D’Agostino et al., 2026).
 525 However, ENSO responses to eruptions are not uniform across reconstructions and models, and depend on eruption magnitude, latitude, seasonality and background climate state (Dee et al., 2020; Zhu et al., 2022; Fang et al., 2025; Wang et al., 2026). This is particularly relevant here because recent work also shows that volcanic eruptions can perturb tropical Atlantic, Pacific and Indo-Pacific coupling, producing spatially contrasted ocean responses rather than a single canonical ENSO anomaly (Verona et al., 2025; Wang et al., 2026). The more useful conclusion is therefore not attribution, but target
 530 identification: these events are the intervals where attribution tests are most likely to be informative.

Proxy composition uncertainty and archive specific contributions add an observing network test to this interpretation. Some features remain visible when tree rings or corals are selectively included or excluded, whereas others change substantially. The 600-700 CE GMT warm optimum is relatively stable across proxy composition experiments, while the 400-600 CE cold episode is more archive sensitive. This contrast suggests that the warm optimum is unlikely to be an artefact of one dominant
 535 archive type, whereas the preceding cold episode may be amplified or damped depending on the relative influence of tree rings, corals and other proxy archives. The 950, 1200 and 1400 CE tropical Pacific and PWC phases persist across proxy composition experiments despite amplitude sensitivity. Their robustness is therefore not only methodological; it is also supported by the archive perturbation experiments.

The archive experiments also clarify why proxy composition matters physically. Tree rings dominate the network and stabilize large scale land temperature information, but their seasonal and moisture sensitivities can influence low frequency
 540 GMT and NASST amplitudes. Corals are sparse and mostly concentrated toward the recent centuries, but their tropical ocean locations give them leverage on marine and circulation related variability. This is consistent with the motivation behind recent realistic pseudoproxy and coral synthesis efforts, which emphasize that sampling, archive processes and proxy system modelling must be tested explicitly rather than treated as generic observational noise (Zhu et al., 2023; Walter et al., 2023).

545 The recent period should be read within the same framework. Industrial era GMT warming is robust in the classical sense: it is product consistent, directly constrained and associated with reduced uncertainty. The positive Niño3.4 tendency is more ambiguous because it is SST based and may include the direct thermal imprint of global warming. The recent negative PWC excursion is more dynamically informative because it is a circulation metric and occurs during a better constrained interval. Together with recent work on disrupted Indo-Pacific coupling, this suggests that recent circulation changes require a longer
 550 paleoclimate context before they can be separated into forced, internally generated and method dependent components (Wang et al., 2026).



Future work should therefore move from uncertainty description to mechanism testing. The high amplitude tropical Pacific and PWC phases identified here should be compared with updated volcanic forcing reconstructions, including eruption timing, latitude and aerosol loading, but the analysis should also allow for internally generated coupled variability and state dependent responses. Superposed epoch analyses, perfect model and pseudoproxy experiments, alternative priors, and nonlinear or state dependent emulators are needed to test whether similar uncertainty reductions occur when the true mechanism is known. New well dated and seasonally interpretable marine archives from the tropical Pacific, Indian Ocean, Maritime Continent, Southern Hemisphere and extratropical ocean basins would be especially valuable because they target the regions that currently control the ambiguity of PWC and tropical Pacific reconstructions. Formal plume distance or related trajectory based comparisons could also be used to test whether our online ensemble, offline PDA products and forced model ensembles are statistically compatible during the high amplitude tropical Pacific and PWC phases.

5 Conclusions

We produced annually resolved Common Era reconstructions of surface air temperature, sea surface temperature, precipitation, evaporation and sea level pressure using an online SIR particle filter coupled to LIM priors trained from CMIP5 and CMIP6 simulations. The experimental design separates posterior internal uncertainty, structural uncertainty from PSM calibration, prior choice and bias correction, and proxy composition uncertainty from tree ring and coral inclusion. The online formulation is also relevant for temporal interpretation: it provides sequentially connected particle histories, while recent work has shown that offline DA plumes require care when used for trajectory level comparisons (Emile-Geay et al., 2025).

The results show that the AllProxy reconstruction contains different kinds of information depending on the variable and period. GMT and NASST support interpretation of the timing and relative prominence of major low frequency features, but their amplitudes remain sensitive to PSM calibration. For GMT, the 600–700 CE warm optimum is relatively stable across proxy composition experiments, whereas the 400–600 CE cold episode is more archive sensitive. This contrast suggests that the warm optimum is unlikely to be an artefact of one dominant archive type, whereas the preceding cold episode may be amplified or damped depending on the relative influence of tree rings, corals and other proxy archives. The early CE cold to warm transition is therefore a meaningful feature of the reconstruction, but the amplitude and physical interpretation of the cold phase require more caution than those of the subsequent warm optimum. NASST shows a broad early CE to late LIA contrast followed by recent warming, with amplitude sensitivity to calibration and coral inclusion.

The strongest dynamical result concerns the tropical Pacific and Pacific Walker circulation. Around approximately 950, 1200 and 1400 CE, Niño3.4, TPDV and PWC show high amplitude phases that coincide with reduced internal and structural uncertainty and persist across proxy composition experiments. These intervals are among the most informative tropical circulation features of the reconstruction. They may reflect volcanic forcing, internally generated coupled ocean atmosphere variability, or interactions between the two, but the present study identifies rather than attributes them.



The recent negative PWC excursion near 2000 CE also stands out because it is a circulation signal rather than a direct SST warming signal and occurs during a comparatively well constrained period. Together with recent work on disrupted Indo-Pacific coupling (Wang et al., 2026), this result suggests that recent tropical circulation changes need to be evaluated against a longer range of CE variability before they can be separated into forced, internally generated and method dependent components.

Future CE reconstructions should combine multi configuration PDA ensembles with explicit archive sensitivity diagnostics and mechanism oriented tests. Reducing uncertainty will require not only more records, but better targeted and seasonally interpretable archives in tropical oceans, the Indian Ocean, the Maritime Continent, the Southern Hemisphere and extratropical ocean basins. These records are needed to determine whether the robust tropical Pacific and PWC phases identified here represent volcanic responses, internally generated dynamical states, or limits of the present proxy prior geometry.



620

Appendix A: Particle filter implementation

A1 State estimation and particle weights

The particle filter is a Bayesian filter that uses a large ensemble of particles to recursively estimate the probability density as a sum of weighted particles:

$$625 \quad p(x_k | y_{1:k}) \approx \sum_{i=1}^N w_k^i \delta(x_k - x_k^i). \quad (A1)$$

Here, x_k is the desired field at time k , $y_{1:k}$ represents the observations up to time k , $\delta(\cdot)$ is a delta function centred on each particle, and w_k^i is the weight associated with particle i . At each annual assimilation step, particles are compared with the available observations and their likelihood is estimated from the following normalized weight:

$$w_{k+1}^i = \frac{\exp\left[-\frac{1}{2}(y_{k+1} - T(x_{k+1}^i))^T C^{-1}(y_{k+1} - T(x_{k+1}^i))\right]}{\sum_{j=1}^N \exp\left[-\frac{1}{2}(y_{k+1} - T(x_{k+1}^j))^T C^{-1}(y_{k+1} - T(x_{k+1}^j))\right]}. \quad (A2)$$

630 In Eq. (A2), C is the error covariance matrix and T is the observation operator that maps the model state onto the space of the available observations.

A2 Error covariance matrix

Because the particle set is propagated in a reduced EOF space, some components of the original fields are not represented.

635 To account for this source of uncertainty, the error covariance matrix is defined as the sum of two independent terms:

$$C_{\text{err}} = C_{\text{obs}} + \sum_{v \in \{T, P\}} |\beta_{v,i}| C_r^v. \quad (A3)$$

Here, C_{obs} is a diagonal matrix containing the PSM calibration residual errors, and C_r is the representativeness error estimated from the mean square difference between the original field and its EOF projection. For each variable v used in the PSM, this error is weighted by the corresponding regression slope.

640

A3 Resampling

The set of particles is resampled according to the weights in Eq. (A2) using the residual resampling approach of Liu and Chen (1998). The integer part of the expected resampling is first allocated to each particle, and the remaining particles are drawn randomly according to the distribution defined by the residual decimal weights. When several proxies are co located at the same grid point, only the proxy record with the highest calibration correlation is retained, which avoids artificially overweighing signals represented by several colocated proxies.

645



650

Appendix B: Construction of the Linear Inverse Models

The LIM emulates the dynamics of each CMIP class model using the following linear stochastic model:

$$\frac{dx}{dt} = Bx + \zeta, \quad (\text{B1})$$

655 where B is a linear operator and x is a Gaussian random forcing term with covariance matrix Q . The fluctuation dissipation relation is written as

$$Q + BC(0) + C(0)B^T = 0, \quad (\text{B2})$$

660 where $C(0)$ is the covariance matrix at lag $t=0$. The B and Q matrices are estimated from long pre existing simulations to reproduce the dynamics simulated by their parent models. Following Penland and Matrosova (1994), a two step second order Runge Kutta method is used to integrate the LIM equation.

Several LIMs are built using last millennium simulations covering 850-1850 CE, including four CMIP5/PMIP3 simulations and four CMIP6/PMIP4 simulations (Schmidt et al., 2011; Taylor et al., 2012; Jungclaus et al., 2017; Kageyama et al., 2018). From these simulations, we derive eight single model LIMs from the leading EOFs of each model. We also build two multi model LIMs by concatenating simulations within each CMIP generation, referred to as LIM CMIP5 and LIM CMIP6.
 665 The EOF analysis is based on five annual mean and linearly detrended variables: surface air temperature at 2 m (Tas), precipitation (Precip), total evaporation (Evap), sea surface temperature (Tos), and sea level pressure (SLP). The number of EOFs retained for each LIM is selected to minimize the number and percentage of negative eigenvalues in Q while maximizing the number of retained modes.

670 **Table B. CMIP models used to construct the LIM priors and number of EOFs retained.**

Model used	Reference	Number of EOFs retained for the LIM without bias correction	Number of EOFs retained for the LIM with bias correction
IPSL-CM6	Boucher et al., 2020; Lurton et al., 2020	22	30
MIROC-ES2L	Ohgaito et al., 2021	29	32



MRI	Yukimoto et al., 2019	20	28
MPI-CM6	Mauritsen et al., 2019; van Dijk et al., 2022	24	20
GISS	Miller et al., 2014	28	23
MIROC	Watanabe et al., 2011	20	30
MPI-CM5	Jungclaus et al., 2013	22	22
NCAR	Gent et al., 2011	24	24
MM-CMIP6 (IPSL-CM6, MIROC- ES2L, MRI, MPI-CM6)		20	22
MM-CMIP5 (GISS, MIROC, MPI- CM5, NCAR)		20	20

675

680

685



Appendix C: Bias correction method

PDA approaches use climate model physics to provide dynamically consistent reconstructions in time and space. In the present framework, the LIM prior propagates information from proxy locations to unsampled regions and variables through the covariance and teleconnection structure inherited from its parent climate model. However, climate models can exhibit systematic spatial biases in mean state, variability amplitude and teleconnection patterns, especially in the tropical oceans. Such biases may affect the way proxy information is propagated through the reconstructed state and may therefore influence regional climate signals, circulation indices and tropical Pacific variability. Spatial pattern bias correction is therefore tested explicitly as one component of the structural uncertainty ensemble, rather than being imposed as a fixed preprocessing choice.

The correction is designed to adjust model spatial patterns associated with selected large scale climate modes toward those estimated from an observationally constrained target. It uses a supervised neural network decoder implemented in Keras (Chollet, 2015). The approach is inspired by encoder decoder architectures commonly used for nonlinear dimensional reduction and reconstruction, but differs from a standard autoencoder because the latent space is not learned freely. Instead, the latent predictors are prescribed climate indices chosen for their relevance to the main modes of variability considered in the reconstruction. This constraint reduces the flexibility of the correction but makes the corrected components physically interpretable.

For each parent model and for the observational target, a decoder is trained to reconstruct annual anomaly fields from a vector of climate indices. The input vector includes the 40 year running mean global mean surface temperature over the 1850-2014 period, used to represent the large scale forced component, together with detrended ENSO, AMV, TPDV and Atl Niño3 indices. TPDV is represented using the tripole Pacific index following Henley et al. (2015). The decoder therefore learns the spatial anomaly patterns of surface air temperature, precipitation and evaporation associated with these indices. In practice, the trained model specific decoder provides the relationship between the prescribed indices and the spatial fields in each parent simulation, while the observational decoder provides the corresponding relationship in the observational target.

The correction is applied in pattern space. For a given model state, the relevant climate indices are first computed and provided as input to both the model specific decoder and the observational decoder. The difference between the observationally based decoded pattern and the model decoded pattern provides an estimate of the spatial pattern adjustment associated with the prescribed climate modes. This adjustment is then added to the original model anomaly field. In schematic form, for a variable Y , model m , year t , and index vector $z_m(t)$, the corrected anomaly field is written as:

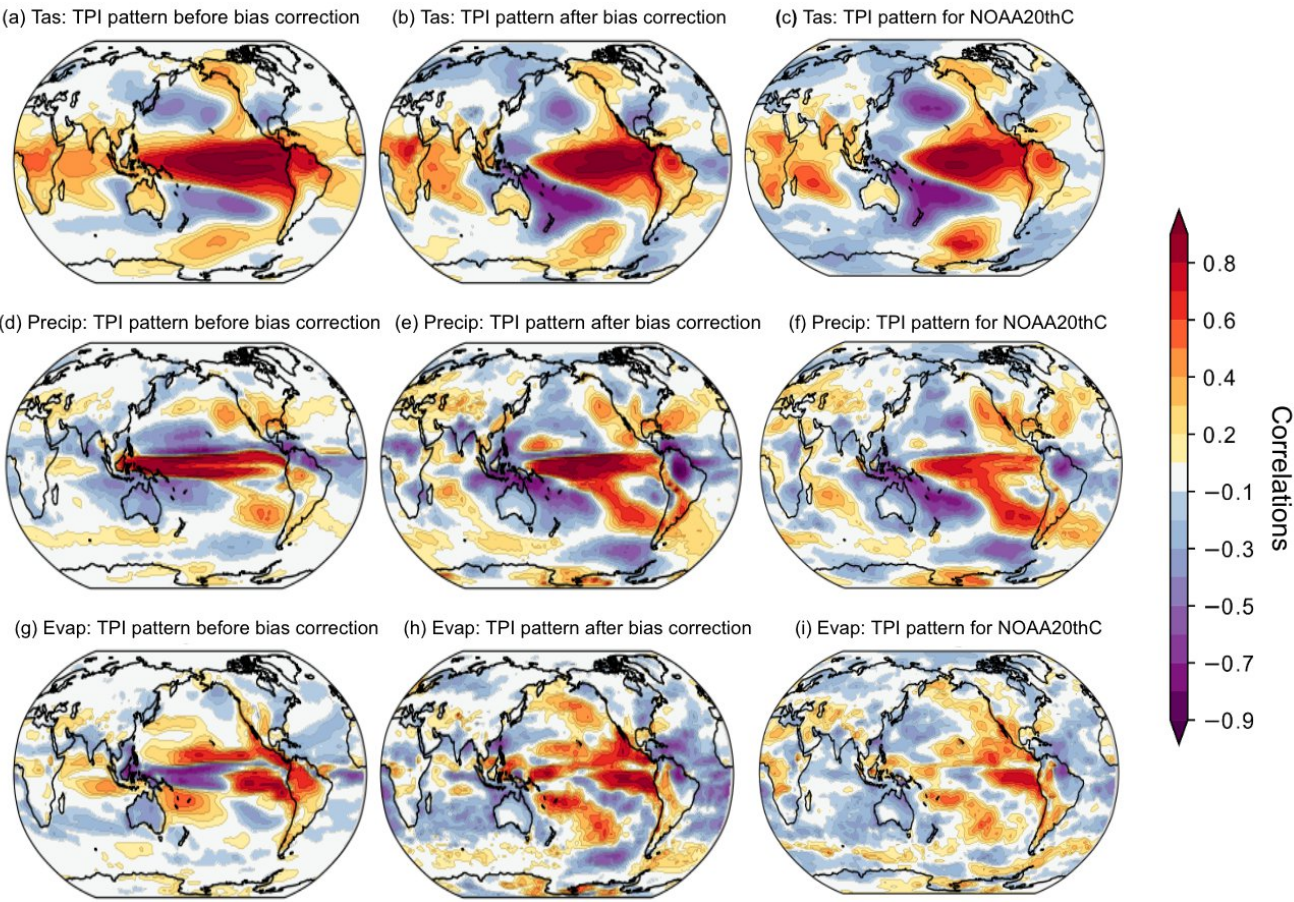
$$Y_m^{\text{corr}}(t) = Y_m(t) + [D_{\text{obs}}(z_m(t)) - D_m(z_m(t))] \quad (\text{C1})$$



Here, $Y_m(t)$ is the original model anomaly field, D_m is the decoder trained on model m , and D_{obs} is the decoder trained on the observational target. The correction therefore modifies the spatial expression of the selected modes while retaining the original temporal evolution of the model indices. It is not intended to correct all aspects of the model state, nor to replace the dynamical prior. Rather, it provides a controlled sensitivity test of how spatial teleconnection biases affect the PDA reconstructions.

Because the observational decoder is trained over the historical period whereas the correction is applied to last millennium simulations, the anthropogenic forcing component is not applied during the last millennium correction. The 40 year running mean global mean surface temperature predictor is used during training to separate the historical forced spatial pattern from natural variability, but this component is set to zero when correcting last millennium simulations. The correction applied to the LIM priors therefore targets only the natural variability patterns associated with ENSO, AMV, TPDV and Atl Niño3. The observational target is NOAA 20th Century Reanalysis version 3 (NOAA20CRv3; Slivinski et al., 2019; Compo et al., 2011), because the correction requires spatially complete and dynamically consistent annual fields of surface air temperature, precipitation and evaporation. NOAA20CRv3 assimilates surface pressure observations and provides a physically consistent multivariate reanalysis suitable for estimating large scale teleconnection patterns over the historical period. The same anomaly definition, spatial grid and temporal filtering are used for the model simulations and the observational target before training the decoders.

Figure C illustrates the correction for the TPDV/TPI teleconnection pattern in the IPSL CM6 last millennium simulation. It compares correlations between the TPI index and surface air temperature, precipitation and evaporation anomaly fields before correction, after correction, and in NOAA20CRv3 over the calibration period covering 1850-2014. This comparison is used as a diagnostic of whether the correction moves the model teleconnection pattern toward the observational target. The figure should be interpreted as a pattern correction diagnostic, not as an independent validation of the full PDA reconstruction.



745 **Figure C: Correlation maps between the TPI index and annual anomaly fields of surface air temperature (a-c), precipitation (d-f) and evaporation (g-i). Maps are shown for the IPSL-CM6A-LR last millennium simulation before bias correction (left column), after bias correction (middle column), and for the NOAA 20th Century Reanalysis version 3 observational target (right column) over the 1850-2014 calibration period. The figure illustrates whether the spatial pattern correction moves the model teleconnection structure toward the observational target.**

750

755



Appendix D: Proxy System Model formulation

760 Different PSMs are calibrated over the instrumental period according to proxy type. For tree ring records, three candidate regression models are considered:

$$\text{Proxy}_i(t) = \begin{cases} \alpha_i + \beta_{T,i} T_{\text{obs}}(t) + \varepsilon_i(t), \\ \alpha_i + \beta_{P,i} P_{\text{obs}}(t) + \varepsilon_i(t), \\ \alpha_i + \beta_{T,i} T_{\text{obs}}(t) + \beta_{P,i} P_{\text{obs}}(t) + \varepsilon_i(t). \end{cases} \quad (\text{D1})$$

For other proxy records, a univariate temperature PSM is used:

$$\text{Proxy}_i(t) = \alpha_i + \beta_{T,i} T_{\text{obs}}(t) + \varepsilon_i(t). \quad (\text{D2})$$

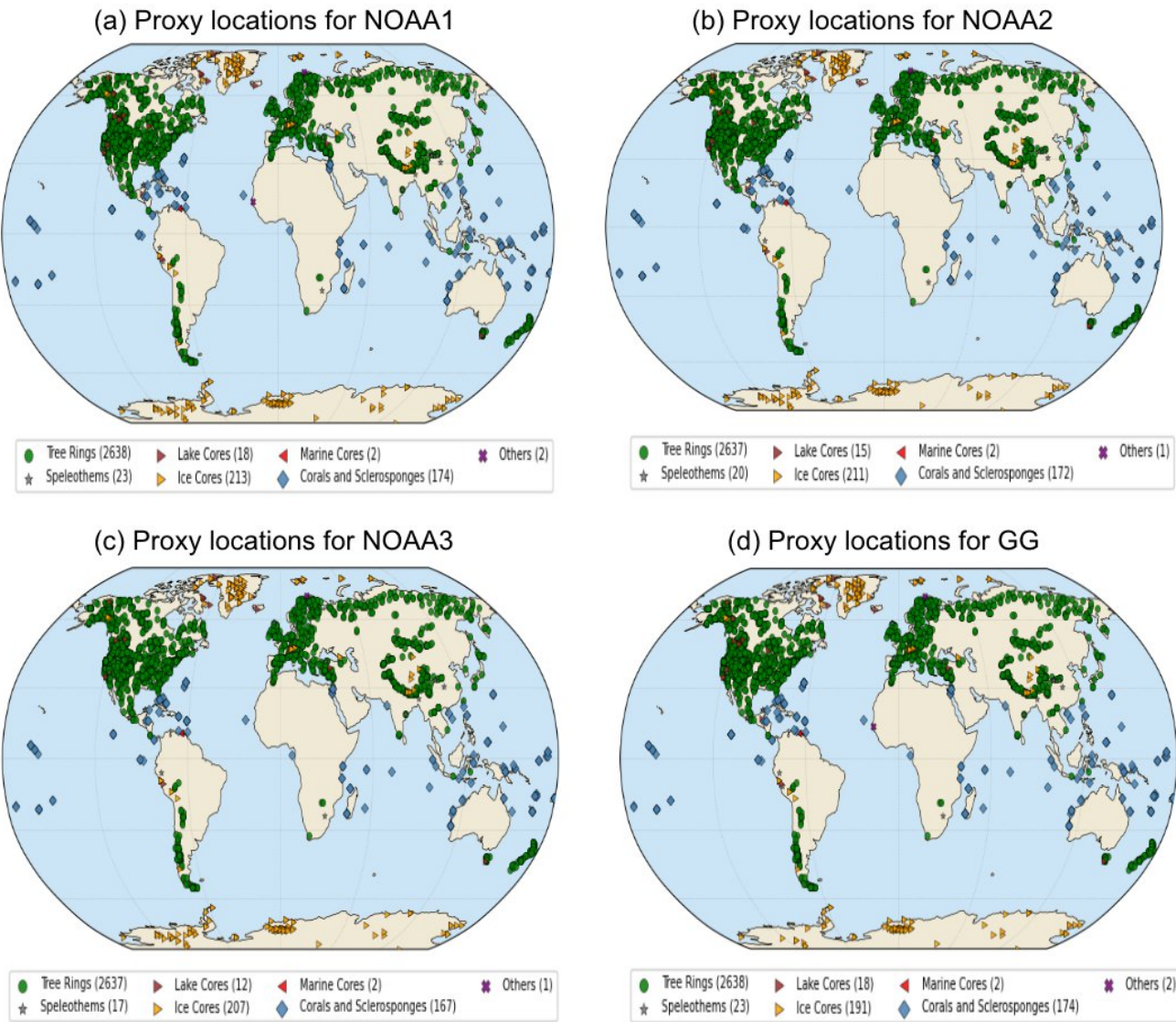
765 Here, α_i is the intercept, $\beta_{T,i}$ and $\beta_{P,i}$ are regression slopes between the standardized proxy series and the instrumental temperature and precipitation time series at the nearest grid point, and $\varepsilon_i(t)$ is the residual error term. For tree ring records, the selected formulation is chosen among the three candidate models using the Bayesian Information Criterion:

$$\text{BIC} = n \ln \left(\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} \right) + k \ln(n) + n(1 + \ln(2\pi)), \quad (\text{D3})$$

where n is the number of years used for calibration and k is the number of estimated parameters.



770 **Appendix E: Proxy database according to the calibration setup**



775 **Figure E: Location and number of proxies used for (a) NOAA1, (b) NOAA2, (c) NOAA3 and (d) GG experiments after removing proxies with less than 20 years of overlap with observations and removing duplicates. The number of assimilated proxies decreases for shorter calibration periods and for datasets with incomplete spatial coverage such as GG.**



Appendix F: Proxy contribution over the instrumental period

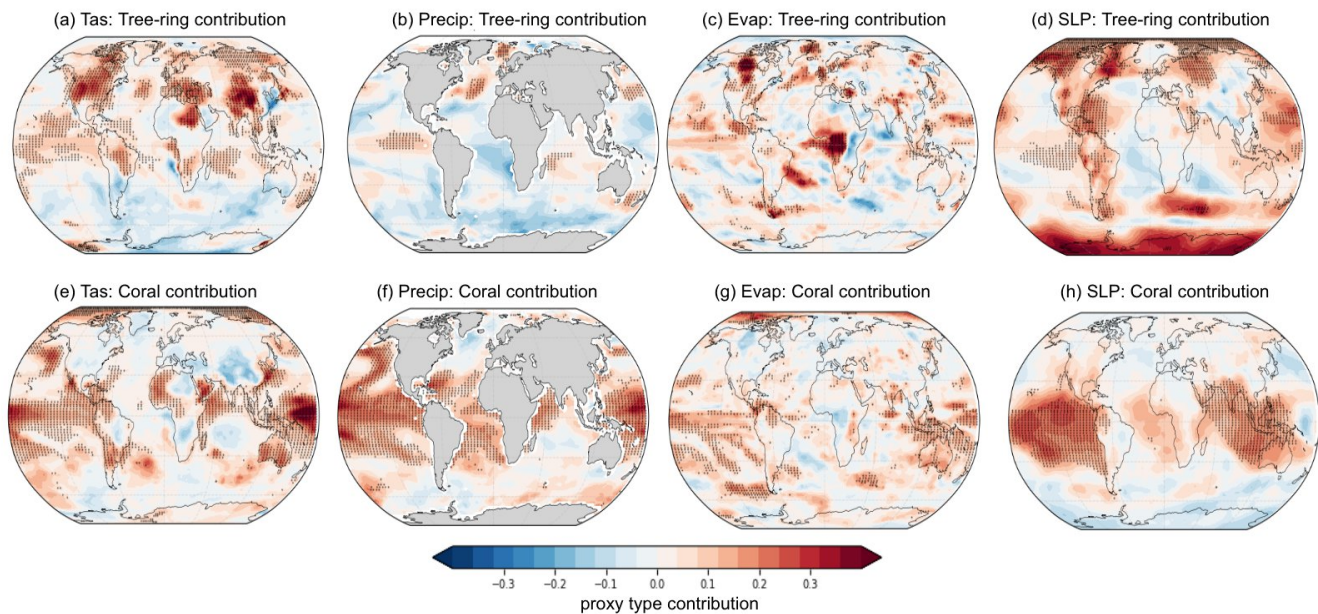


Figure F: Proxy contribution over the 1900-2000 period. Each map shows the difference in correlation with observations (ΔCorr) between the AllProxy median reconstruction and the median reconstruction when excluding tree rings (first row) and corals (second row) for: surface air temperature (Tas), sea surface temperature (Tos), precipitation (Pr), and sea-level pressure (SLP). Positive values (red) indicate that including the proxy type improves the reconstruction skill locally, and negative values (blue) indicate a degradation. Stippling marks regions where all priors agree on the sign.



Appendix G: Comparison with existing products

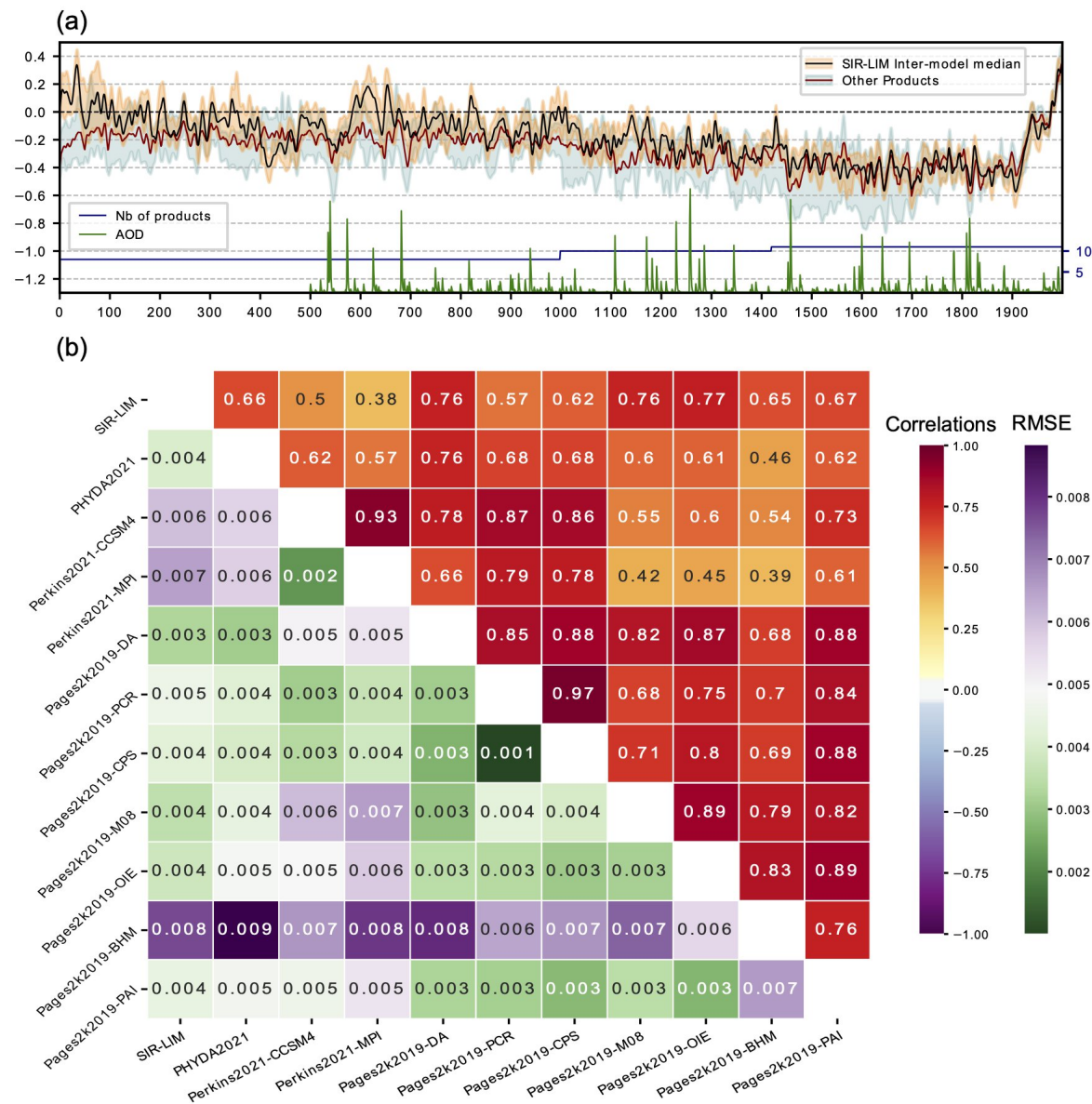


Figure G: Comparison of the NOAA1 global mean surface temperature reconstruction with existing last millennium reconstructions and instrumental products. (a) Reconstructed global mean surface temperature (T_{as} , °C) from the NOAA1 reference experiment, compared with independent last millennium reconstructions and instrumental products. The black line shows NOAA1 inter model median, and the orange shading indicates the ensemble spread. Coloured lines show Pages2k2019 DA, PHYDA2021, Perkins2021-MPI, Perkins2021-CCSM4, ModERA and NOAA20CRv3. The green curve shows global mean stratospheric aerosol optical depth (AOD; Toohey and Sigl, 2017), used as an indicator of major volcanic eruptions. (b) Pairwise comparison of global mean surface temperature time series over 1421-2000 CE. The upper triangle shows Pearson correlation coefficients between each pair of products, and the lower triangle shows the corresponding root mean square error (RMSE), with colour scales indicated on the right.



Appendix H: Spatial coherence across methodological configurations

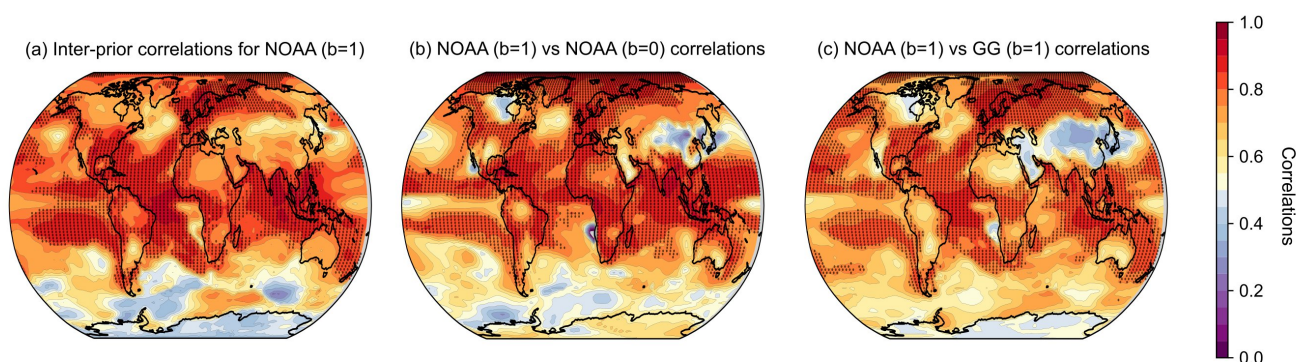


Figure H: Spatial coherence across methodological configurations. Spatial correlations between reconstructed fields across the same methodological comparisons as in Fig. 6. (a) Inter prior consistency within the NOAA1($b=1$) experiment. (b) Effect of bias correction, comparing NOAA1($b=1$) and NOAA1($b=0$). (c) Effect of PSM calibration, comparing NOAA1($b=1$) and GG($b=1$). Stippling indicates regions where at least 90 % of the comparisons exceed the selected correlation threshold.



Appendix I: Scale dependence of structural uncertainty for GMT, NASST and TPDV

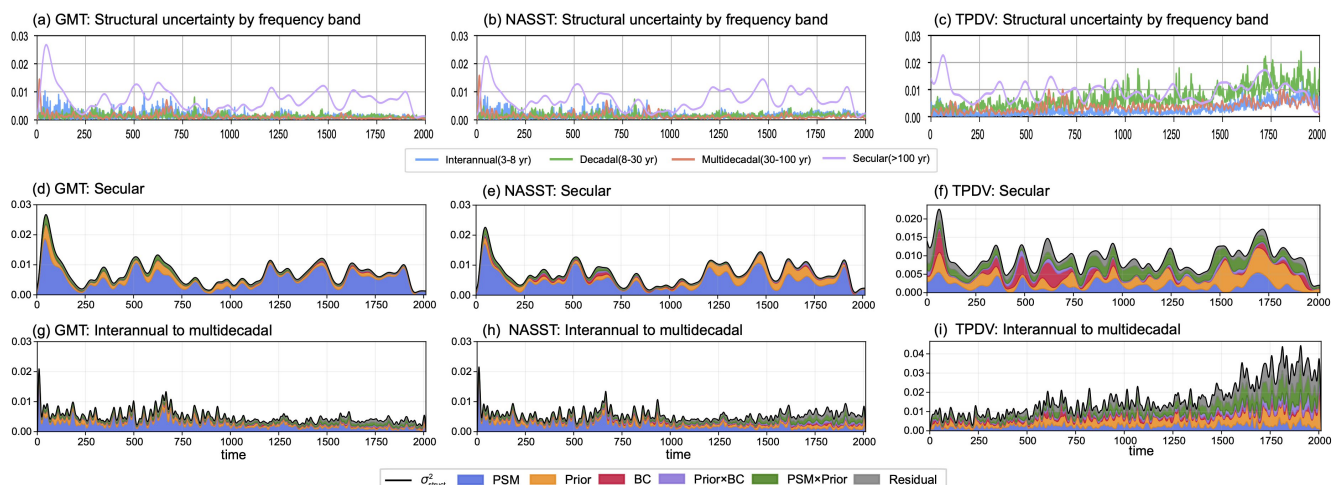
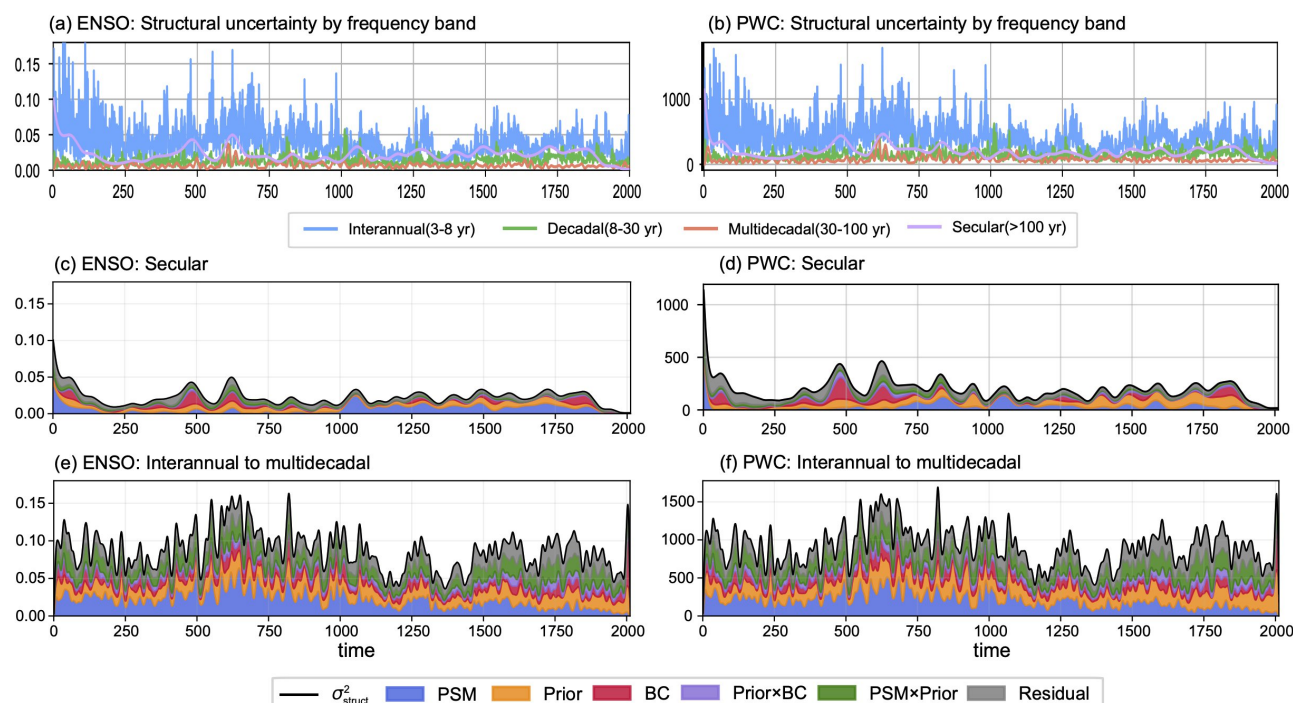


Figure I: Scale dependence of structural uncertainty for GMT, NASST and TPDV. Structural uncertainty across the 80 configurations, which differ in PSM calibration, LIM prior and bias correction, is shown for global mean surface temperature (GMT; a, d, g, j), North Atlantic sea surface temperature (NASST; b, e, h, k) and tropical Pacific decadal variability (TPDV; c, f, i, l). Panels (a-c) show the full structural uncertainty decomposition. Panels (d-f) show the decomposition of total structural variance by frequency band: interannual (3-8 years), decadal (8-30 years), multidecadal (30-100 years) and secular (>100 years). Panels (g-i) show the ANOVA decomposition of structural uncertainty for secular variability (>100 years), with the relative contributions of PSM calibration, prior choice, bias correction (BC) and their interactions. Panels (j-l) show the same decomposition for interannual to multidecadal variability (<100 years).



Appendix J: Scale dependence of structural uncertainty for Niño3.4 and PWC



855 **Figure J: Scale dependence of structural uncertainty for Niño3.4 and PWC. Same as Fig. I, but for the Niño3.4 index (a, c, e) and**
 the Pacific Walker circulation (PWC) index (b, d, f). Panels (a, b) show the decomposition of total structural variance by frequency
 band: interannual (3-8 years), decadal (8-30 years), multidecadal (30-100 years) and secular (>100 years). Panels (c, d) show the
 ANOVA decomposition of structural uncertainty for secular variability (>100 years), and panels (e, f) show the same
 860 decomposition for interannual to multidecadal variability (<100 years), with the relative contributions of PSM calibration, prior
 choice, bias correction (BC) and their interactions.



Code and data availability

The reconstruction ensemble, derived indices, uncertainty decompositions, figure generation scripts and the processed proxy data used as input are archived on Zenodo (DOI: 10.5281/zenodo.22690510). The original proxy databases can also be retrieved from their respective repositories: the PAGES2k v2.0.0 temperature database (PAGES2k Consortium, 2017; <https://doi.org/10.6084/m9.figshare.c.3285353>), the Iso2kv1.0.0 database (Konecky et al., 2020; http://lipdverse.org/iso2k/current_version/) and the proxy compilation distributed with the PHYDA reconstruction (Steiger et al., 2018; <https://www.ncei.noaa.gov/pub/data/paleo/reconstructions/steiger2018/>).

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

870 Author contributions

B.J. developed and implemented the reconstruction experiments, performed the sensitivity analyses, produced the figures, prepared an initial draft of selected sections, and contributed to the analysis of the results. M.K. conceived the study, designed the uncertainty quantification framework, supervised the work, led the scientific interpretation of the results, substantially restructured the manuscript, and wrote the final version. B.J. and M.K. discussed the results, revised the manuscript and approved the final version.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Authors acknowledge support from the IPSL Climate Graduate School EUR and from the HPC resources of TGCC under grant nos. A0150113826, A0170113826 and A0190113826 provided by GENCI (Grand Équipement National de Calcul Intensif). This study benefited from the ESPRI (Ensemble de Services Pour la Recherche l'IPSL) computing and data centre (<https://mesocentre.ipsl.fr>, last access: 30 July 2026) which is supported by CNRS, Sorbonne Université, École Polytechnique and CNES.



Financial support

885 This work received government funding managed by the French National Research Agency under France 2030, as part of the TRACCS ARCHIVES project, reference ANR 25 EXTR 0001.

References

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present), *J. Hydrometeorol.*, 4, 1147-1167, [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2), 2003.
- Allan, R. and Ansell, T.: A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850-2004, *J. Climate*, 19, 5816-5842, <https://doi.org/10.1175/JCLI3937.1>, 2006.
- Amrhein, D. E., Hakim, G. J., and Parsons, L. A.: Quantifying structural uncertainty in paleoclimate data assimilation with an application to the Last Millennium, *Geophys. Res. Lett.*, 47, e2020GL090485, <https://doi.org/10.1029/2020GL090485>, 2020.
- 890 Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M.: A description of the global land surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial trend analysis from 1901-present, *Earth Syst. Sci. Data*, 5, 71-99, [https://doi.org/10.5194/essd-5-71-99-](https://doi.org/10.5194/essd-5-71-99-2013)
- 900 2013, 2013.
- Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Ethé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, L., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and evaluation of the
- 910 IPSL-CM6A-LR climate model, *J. Adv. Model. Earth Sy.*, 12, e2019MS002010, <https://doi.org/10.1029/2019MS002010>, 2020.
- Breitenmoser, P., Brönnimann, S., and Frank, D.: Forward modelling of tree ring width and comparison with a global network of tree ring chronologies, *Clim. Past*, 10, 437-449, <https://doi.org/10.5194/cp-10-437-2014>, 2014.
- Chollet, F.: Keras, GitHub repository [code], <https://github.com/keras-team/keras>, 2015.



- 915 Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, *Q. J. Roy. Meteor. Soc.*, 137, 1-28, <https://doi.org/10.1002/qj.776>, 2011.
- 920 D'Agostino, R., Fang, S.-W., Khodri, M., Pausata, F. S. R., Tejedor, E., Timmreck, C., and Zanchettin, D.: The Impact of Volcanic Eruptions on Tropical Hydroclimate, *Annu. Rev. Earth Pl. Sc.*, 54, 365-387, <https://doi.org/10.1146/annurev-earth-071025-112048>, 2026.
- Dee, S. G., Cobb, K. M., Emile-Geay, J., Ault, T. R., Edwards, R. L., Cheng, H., and Charles, C. D.: No consistent ENSO response to volcanic forcing over the last millennium, *Science*, 367, 1477-1481, <https://doi.org/10.1126/science.aax2000>, 925 2020.
- Emile-Geay, J., Hakim, G. J., Viens, F., Zhu, F., and Amrhein, D. E.: Temporal comparisons involving paleoclimate data assimilation: challenges and remedies, *J. Climate*, 38, 1365-1385, <https://doi.org/10.1175/JCLI-D-24-0101.1>, 2025.
- Falster, G., Konecky, B., Coats, S., and Stevenson, S.: Forced changes in the Pacific Walker circulation over the past millennium, *Nature*, 622, 93-100, <https://doi.org/10.1038/s41586-023-06447-0>, 2023.
- 930 Fang, S.-W., Pausata, F. S. R., D'Agostino, R., Khodri, M., Zanchettin, D., and Timmreck, C.: Seasonal ITCZ Control on ENSO Responses to Extratropical Volcanic Forcing, *J. Climate*, 38, 6061-6079, <https://doi.org/10.1175/JCLI-D-24-0554.1>, 2025.
- Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., and Zhang, M.: The Community Climate System Model version 4, 935 *J. Climate*, 24, 4973-4991, <https://doi.org/10.1175/2011JCLI4083.1>, 2011.
- Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, *Rev. Geophys.*, 48, RG4004, <https://doi.org/10.1029/2010RG000345>, 2010.
- Hawkins, E., Smith, R. S., Gregory, J. M., and Stainforth, D. A.: Irreducible uncertainty in near-term climate projections, *Clim. Dynam.*, 46, 3807-3819, <https://doi.org/10.1007/s00382-015-2806-8>, 2016.
- 940 Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., and Folland, C. K.: A Tripole Index for the Interdecadal Pacific Oscillation, *Clim. Dynam.*, 45, 3077-3090, <https://doi.org/10.1007/s00382-015-2525-1>, 2015.
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons, *J. Climate*, 30, 8179-8205, <https://doi.org/10.1175/JCLI-D-16-0836.1>, 2017.
- 945 Jebri, B. and Khodri, M.: Large ensemble particle filter for spatial climate reconstructions using a linear inverse model, *J. Adv. Model. Earth Sy.*, 15, e2022MS003094, <https://doi.org/10.1029/2022MS003094>, 2023.
- Jungclauss, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., Egorova, T., Evans, M., González-Rouco, J. F., Goosse, H., Hurrett, G. C., Joos, F., Kaplan, J. O., Khodri, M., Klein Goldewijk, K., Krivova, N., LeGrande, A. N., Lorenz, S.



- J., Luterbacher, J., Man, W., Maycock, A. C., Meinshausen, M., Moberg, A., Muscheler, R., Nehrbass-Ahles, C., Otto-
950 Bliesner, B. L., Phipps, S. J., Pongratz, J., Rozanov, E., Schmidt, G. A., Schmidt, H., Schmutz, W., Schurer, A., Shapiro, A.
I., Sigl, M., Smerdon, J. E., Solanki, S. K., Timmreck, C., Toohey, M., Usoskin, I. G., Wagner, S., Wu, C.-J., Yeo, K. L.,
Zanchettin, D., Zhang, Q., and Zorita, E.: The PMIP4 contribution to CMIP6, Part 3: the last millennium, scientific
objective, and experimental design for the PMIP4 past1000 simulations, *Geosci. Model Dev.*, 10, 4005-4033,
<https://doi.org/10.5194/gmd-10-4005-2017>, 2017.
- 955 Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch,
J.-S.: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM), the ocean component of
the MPI Earth system model, *J. Adv. Model. Earth Sy.*, 5, 422-446, <https://doi.org/10.1002/jame.20023>, 2013.
- Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Peterschmitt, J.-Y.,
Abe-Ouchi, A., Albani, S., Bartlein, P. J., Brierley, C., Crucifix, M., Dolan, A., Fernandez-Donado, L., Fischer, H.,
960 Hopcroft, P. O., Ivanovic, R. F., Lambert, F., Lunt, D. J., Mahowald, N. M., Peltier, W. R., Phipps, S. J., Roche, D. M.,
Schmidt, G. A., Tarasov, L., Valdes, P. J., Zhang, Q., and Zhou, T.: The PMIP4 contribution to CMIP6, Part 1: overview and
over-arching analysis plan, *Geosci. Model Dev.*, 11, 1033-1057, <https://doi.org/10.5194/gmd-11-1033-2018>, 2018.
- Khodri, M., Izumo, T., Vialard, J., Janicot, S., Cassou, C., Lengaigne, M., Mignot, J., Gastineau, G., Guilyardi, E., Lebas,
N., Robock, A., and McPhaden, M. J.: Tropical explosive volcanic eruptions can trigger El Niño by cooling tropical Africa,
965 *Nat. Commun.*, 8, 778, <https://doi.org/10.1038/s41467-017-00755-6>, 2017.
- Konecky, B. L., McKay, N. P., Churakova (Sidorova), O. V., Comas-Bru, L., Dassié, E. P., DeLong, K. L., Falster, G. M.,
Fischer, M. J., Jones, M. D., Jonkers, L., Kaufman, D. S., Leduc, G., Managave, S. R., Martrat, B., Opel, T., Orsi, A. J.,
Partin, J. W., Sayani, H. R., Thomas, E. K., Thompson, D. M., Tyler, J. J., Abram, N. J., Atwood, A. R., Cartapanis, O.,
Conroy, J. L., Curran, M. A., Dee, S. G., Deininger, M., Divine, D. V., Kern, Z., Porter, T. J., Stevenson, S. L., von Gunten,
970 L., and Iso2k Project Members: The Iso2k database: a global compilation of paleo-d18O and d2H records to aid
understanding of Common Era climate, *Earth Syst. Sci. Data*, 12, 2261-2288, <https://doi.org/10.5194/essd-12-2261-2020>,
2020.
- Laepple, T., Ziegler, E., Weitzel, N., Hébert, R., Ellerhoff, B., Schoch, P., Martrat, B., Bothe, O., Moreno-Chamarro, E.,
Chevalier, M., Herbert, A., and Rehfeld, K.: Regional but not global temperature variability underestimated by climate
975 models at supradeccadal timescales, *Nat. Geosci.*, 16, 958-966, <https://doi.org/10.1038/s41561-023-01299-9>, 2023.
- Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the
GISTEMP uncertainty model, *J. Geophys. Res.-Atmos.*, 124, 6307-6326, <https://doi.org/10.1029/2018JD029522>, 2019.
- Liu, J. S. and Chen, R.: Sequential Monte Carlo methods for dynamic systems, *J. Am. Stat. Assoc.*, 93, 1032-1044,
<https://doi.org/10.1080/01621459.1998.10473765>, 1998.
- 980 Luo, X., Dee, S. G., Okumura, Y. M., and Parsons, L. A.: Nonstationary teleconnections over North America revealed in
paleoclimate data assimilation reconstructions spanning the Last Millennium, *J. Geophys. Res.-Atmos.*, 129,
e2024JD041150, <https://doi.org/10.1029/2024JD041150>, 2024.



- Lurton, T., Balkanski, Y., Bastrikov, V., Bekki, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Contoux, C., Cozic, A., Cugnet, D., Dufresne, J.-L., Ethé, C., Foujols, M.-A., Ghattas, J., Hauglustaine, D., Hu, R.-M., Kageyama, M., Khodri, M., Lebas, N., Levavasseur, G., Marchand, M., Ottlé, C., Peylin, P., Sima, A., Szopa, S., Thiéblemont, R., Vuichard, N., and Boucher, O.: Implementation of the CMIP6 forcing data in the IPSL-CM6A-LR model, *J. Adv. Model. Earth Sy.*, 12, e2019MS001940, <https://doi.org/10.1029/2019MS001940>, 2020.
- Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jiménéz-de-la-Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kornblueh, L., Marotzke, J., Matei, D., Mikolajewicz, U., Müller, W. A., Notz, D., Pongratz, J., Rast, S., Redler, R., Reick, C. H., Schmidt, H., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Wieners, K.-H., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO₂, *J. Adv. Model. Earth Sy.*, 11, 998-1038, <https://doi.org/10.1029/2018MS001400>, 2019.
- McPhaden, M. J., Zebiak, S. E., and Glantz, M. H.: ENSO as an integrating concept in Earth science, *Science*, 314, 1740-1745, <https://doi.org/10.1126/science.1132588>, 2006.
- Miller, R. L., Schmidt, G. A., Nazarenko, L. S., Tausnev, N., Bauer, S. E., Del Genio, A. D., Kelley, M., Lo, K. K., Ruedy, R., Shindell, D. T., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Chen, Y.-H., Cheng, Y., Clune, T. L., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Lo, K. K., Menon, S., Oinas, V., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Russell, G. L., Sato, M., Sun, S., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: CMIP5 historical simulations (1850-2012) with GISS ModelE2, *J. Adv. Model. Earth Sy.*, 6, 441-477, <https://doi.org/10.1002/2013MS000266>, 2014.
- Ohgaito, R., Yamamoto, A., Hajima, T., O'ishi, R., Abe, M., Tatebe, H., Abe-Ouchi, A., and Kawamiya, M.: PMIP4 experiments using MIROC-ES2L Earth system model, *Geosci. Model Dev.*, 14, 1195-1217, <https://doi.org/10.5194/gmd-14-1195-2021>, 2021.
- PAGES 2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, *Sci. Data*, 4, 170088, <https://doi.org/10.1038/sdata.2017.88>, 2017.
- PAGES 2k Consortium: Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era, *Nat. Geosci.*, 12, 643-649, <https://doi.org/10.1038/s41561-019-0400-0>, 2019.
- PAGES Hydro2k Consortium: Comparing proxy and model estimates of hydroclimate variability and change over the Common Era, *Clim. Past*, 13, 1851-1900, <https://doi.org/10.5194/cp-13-1851-2017>, 2017.
- Penland, C. and Matrosova, L.: A balance condition for stochastic numerical models with application to the El Niño-Southern Oscillation, *J. Climate*, 7, 1352-1372, [https://doi.org/10.1175/1520-0442\(1994\)007<1352:ABCFSN>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1352:ABCFSN>2.0.CO;2), 1994.
- Perkins, W. A. and Hakim, G. J.: Coupled atmosphere-ocean reconstruction of the Last Millennium using online data assimilation, *Paleoceanogr. Paleocl.*, 36, e2020PA003959, <https://doi.org/10.1029/2020PA003959>, 2021.



- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.-Atmos.*, 108, 4407, <https://doi.org/10.1029/2002JD002670>, 2003.
- 1020 Rohde, R. A. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, *Earth Syst. Sci. Data*, 12, 3469-3479, <https://doi.org/10.5194/essd-12-3469-2020>, 2020.
- Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), *Geosci. Model Dev.*, 4, 33-45, <https://doi.org/10.5194/gmd-4-33-2011>, 2011.
- 1025 Schneider, U., Hänsel, S., Finger, P., Rustemeier, E., and Ziese, M.: GPCC Full Data Monthly Version 2022 at 0.25°: monthly land surface precipitation from rain gauges built on GTS based and historic data, *Deutscher Wetterdienst*, https://doi.org/10.5676/DWD_GPCC/FD_M_V2022_025, 2022.
- Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Le Blancq, F., Lee, T. C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., Wyszynski, P., and Worley, S. J.: Towards a more reliable historical reanalysis: improvements for version 3 of the Twentieth Century Reanalysis system, *Q. J. Roy. Meteor. Soc.*, 145, 2876-2908, <https://doi.org/10.1002/qj.3598>, 2019.
- 1030 Steiger, N. J., Smerdon, J. E., Cook, E. R., and Cook, B. I.: A reconstruction of global hydroclimate and dynamical variables over the Common Era, *Sci. Data*, 5, 180086, <https://doi.org/10.1038/sdata.2018.86>, 2018.
- Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling, *Clim. Past*, 15, 1251-1273, <https://doi.org/10.5194/cp-15-1251-2019>, 2019.
- 1040 Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *B. Am. Meteorol. Soc.*, 93, 485-498, <https://doi.org/10.1175/BAMS-D-11-00094.1>, 2012.
- Toohey, M. and Sigl, M.: Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE, *Earth Syst. Sci. Data*, 9, 809-831, <https://doi.org/10.5194/essd-9-809-2017>, 2017.
- 1045 Valler, V., Franke, J., Brugnara, Y., Samakinwa, E., Hand, R., Lundstad, E., Burgdorf, A.-M., Lipfert, L., Friedman, A. R., and Brönnimann, S.: ModE-RA: a global monthly paleo-reanalysis of the modern era 1421 to 2008, *Sci. Data*, 11, 36, <https://doi.org/10.1038/s41597-023-02733-8>, 2024.



- van Dijk, E., Jungclauss, J., Lorenz, S., Timmreck, C., and Krüger, K.: Was there a volcanic-induced long-lasting cooling
 1050 over the Northern Hemisphere in the mid-6th-7th century?, *Clim. Past*, 18, 1601-1623, [https://doi.org/10.5194/cp-18-1601-](https://doi.org/10.5194/cp-18-1601-2022)
 2022, 2022.
- Verona, L., Khodri, M., and Wainer, I.: Out-of-phase tropical Atlantic-Pacific response after volcanic eruptions over the Last
 Millennium in PMIP4/past1000 simulations, *Sci. Rep.*, 15, 45788, <https://doi.org/10.1038/s41598-025-28495-4>, 2025.
- Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M.,
 1055 Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M.: MIROC-ESM 2010: model description
 and basic results of CMIP5-20c3m experiments, *Geosci. Model Dev.*, 4, 845-872, <https://doi.org/10.5194/gmd-4-845-2011>,
 2011.
- Walter, R. M., Sayani, H. R., Felis, T., Cobb, K. M., Abram, N. J., Arzey, A. K., Atwood, A. R., Brenner, L. D., Dassié, E.
 P., DeLong, K. L., and the PAGES CoralHydro2k Project Members: The CoralHydro2k database: a global, actively curated
 1060 compilation of coral $\delta^{18}\text{O}$ and Sr/Ca proxy records of tropical ocean hydrology and temperature for the Common Era, *Earth*
Syst. Sci. Data, 15, 2081-2116, <https://doi.org/10.5194/essd-15-2081-2023>, 2023.
- Wang, S., Oppo, D. W., and Ummenhofer, C. C.: Coupling of Pacific and Indian Ocean variability disrupted by 19th century
 volcanism, *Nat. Commun.*, 17, 8572, <https://doi.org/10.1038/s41467-026-76705-y>, 2026.
- Wu, F., Ning, L., Liu, Z., Liu, J., Hu, W., Yan, M., Xing, F., Lei, L., Sun, H., Chen, K., Qin, Y., Li, B., and Xu, C.: A new
 1065 last two millennium reanalysis based on hybrid gain analog offline EnKF and an expanded proxy database, *npj Clim. Atmos.*
Sci., 8, 62, <https://doi.org/10.1038/s41612-025-00961-w>, 2025.
- Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka,
 M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., and Ishii, M.: The Meteorological Research
 Institute Earth System Model version 2.0, MRI-ESM2.0: description and basic evaluation of the physical component, *J.*
 1070 *Meteorol. Soc. Jpn.*, 97, 931-965, <https://doi.org/10.2151/jmsj.2019-051>, 2019.
- Zhu, F., Emile-Geay, J., Hakim, G. J., King, J., and Anchukaitis, K. J.: A re-appraisal of the ENSO response to volcanism
 with paleoclimate data assimilation, *Nat. Commun.*, 13, 747, <https://doi.org/10.1038/s41467-022-28210-1>, 2022.
- Zhu, F., Emile-Geay, J., Anchukaitis, K. J., McKay, N. P., Stevenson, S., and Meng, Z.: A pseudoproxy emulation of the
 PAGES 2k database using a hierarchy of proxy system models, *Sci. Data*, 10, 624, [https://doi.org/10.1038/s41597-023-](https://doi.org/10.1038/s41597-023-02489-1)
 1075 02489-1, 2023.