Coupled Atmosphere-Ocean Reconstruction of the Last Millennium Using Online Data Assimilation

Walter Perkins[1] and Gregory Hakim[2]

[1]Vulcan
[2]University of Washington

November 23, 2022

Abstract

We use online data assimilation to combine information from a linear inverse model of coupled atmosphere-ocean dynamics with proxy records to create a new annual-resolution reconstruction of atmosphere and ocean fields over the last millennium. Instrumental validation of reconstructed sea-surface temperature and 0-700 m ocean heat content shows broad regions of positive spatial correlations, and high correlations (˜0.6-0.9) for global averages and indices of large-scale modes of atmospheric variability. Compared to previous reconstructions, the online reconstructions show global and hemispheric averages with little-to-no millennial-scale trend and global-mean temperatures ˜0.25-0.5 K cooler during early periods (1000-1400 C.E.). The spatial anomaly differences of average temperature between an early (1000-1250 C.E.) and later (1400-1700 C.E.) period show warm anomalies over high-latitude Europe and cool tropical conditions in partial agreement with previous assessments. The addition of online data assimilation, which provides dynamical memory to climate proxy information, is shown to be crucial for adequately characterizing decadal-to-centennial-scale variability of 0–700 m ocean heat content. Furthermore, the climate forecasts provide model-based physical constraints for atmosphere-ocean interaction, which become increasingly important during early periods when less proxy information is available for assimilation.

Coupled Atmosphere–Ocean Reconstruction of the Last Millennium Using Online Data Assimilation

W. A. Perkins[1], **G. J. Hakim**[2]

[1]Vulcan, Inc, Seattle, WA USA
[2]University of Washington, Seattle, WA USA

Key Points:

- A new online paleoclimate data assimilation method for atmosphere–ocean reconstruction over the last millennium provides dynamical proxy memory
- Reconstructed ocean field validation against instrumental products is largely skillful despite a sparse proxy network
- Online data assimilation improves the dynamics and low-frequency variability of reconstructed coupled fields relative to offline assimilation

Corresponding author: Walter A. Perkins, `wperkins@uw.edu`

Abstract

We use online data assimilation to combine information from a linear inverse model of coupled atmosphere-ocean dynamics with proxy records to create a new annual-resolution reconstruction of atmosphere and ocean fields over the last millennium. Instrumental validation of reconstructed sea-surface temperature and 0–700 m ocean heat content shows broad regions of positive spatial correlations, and high correlations ($\sim$0.6–0.9) for global averages and indices of large-scale modes of atmospheric variability. Compared to previous reconstructions, the online reconstructions show global and hemispheric averages with little-to-no millennial-scale trend and global-mean temperatures $\sim$0.25–0.5 K cooler during early periods (1000–1400 C.E.). The spatial anomaly differences of average temperature between an early (1000–1250 C.E.) and later (1400–1700 C.E.) period show warm anomalies over high-latitude Europe and cool tropical conditions in partial agreement with previous assessments. The addition of online data assimilation, which provides dynamical memory to climate proxy information, is shown to be crucial for adequately characterizing decadal-to-centennial-scale variability of 0–700 m ocean heat content. Furthermore, the climate forecasts provide model-based physical constraints for atmosphere–ocean interaction, which become increasingly important during early periods when less proxy information is available for assimilation.

1 Introduction

Defining the range and mechanisms of low-frequency variability of the Earth system is a crucial factor in understanding climate sensitivity and impacts of future climate change. At least for decadal-to-centennial timescales, the oceans play a central role, acting as an energy reservoir that integrates over the noisier and chaotic atmosphere (e.g., Hasselmann, 1976). While the instrumental record provides a direct account of climate system variability, the length of record is relatively short for investigating slow climate features, especially when considering coupled atmosphere–ocean variability. High-resolution reanalysis products partly extend the record by using numerical models to assimilate observations into atmosphere and ocean fields (e.g., Compo et al., 2011; Balmaseda et al., 2013; Chang et al., 2013; Poli et al., 2016). However, coupled atmosphere–ocean reanalysis products are in the early stage of development (Laloyaux et al., 2018), and reanalysis products are still limited in time due to the lack of observations earlier in the 20th Century. The proxy record, including measurements from trees, ice, corals, sediments, and more, provides an extended account of observed information to investigate decadal and longer climate variability. In order to extend the information contained in proxies from the locations and times to which they pertain, additional information is needed to create a climate field reconstruction (CFR). Early CFR approaches derived the additional information from static statistical relationships derived from the instrumental era (Mann et al., 2009). More recently, paleoclimate data assimilation (PDA) uses dynamical constraints encoded in global climate models (GCMs) to provide the additional information (Hakim et al., 2016; Franke et al., 2017; Okazaki & Yoshimura, 2017; Steiger et al., 2018; Tardif et al., 2019). The CFRs include robust uncertainty quantification and a set of physically consistent spatial fields useful for dynamical inquiry beyond what is feasible from proxies or climate models alone (e.g., Singh et al., 2017). Here we present a new PDA method that uses an approximate GCM to propagate the information extracted from proxies through time, providing dynamical memory, which dramatically improves the reconstruction of low-frequency climate variability. We use these new results to explore low-frequency aspects of coupled atmosphere-ocean dynamics that are not possible with traditional CFRs.

Ensemble-based data assimilation techniques, such as those adapted for PDA, traditionally include a model forecast that translates the updated analysis state to the next time for use as a prior state estimate. This technique, known as "online" assimilation, assumes the forecast model possesses at least some predictive skill at the same timescale as the time interval between assimilated observations (e.g., Pendergrass et al., 2012). In the PDA context, the lack of significant climate forecast skill and computational expense of performing GCM forecasts has led to most studies omitting the use of a forecast model (Huntley & Hakim, 2010; Bhend et al., 2012; Steiger et al., 2014; Hakim et al., 2016; Franke et al., 2017; Steiger et al., 2018; Tardif et al., 2019). The no-forecast approach, known as an "offline" method (Oke, 2002; Evensen, 2003; Oke et al., 2005, 2007), uses climatological information (e.g., from existing climate simulations) to generate a prior estimate of fields for each analysis time. This technique benefits from being computationally efficient, but the state trajectory over time is solely dependent on available proxy information and is not constrained by physics. The oceans are a potential source of predictability for up to decadal timescales (e.g., Hawkins & Sutton, 2009; Branstator et al., 2012; Zanna, 2012) due to memory and predictable dynamics (e.g., the El Niño Southern Oscillation; ENSO). Thus, the incorporation of a forecast model encapsulating these processes could help constrain low-frequency variability that may not be well represented by proxies (Broecker, 2001; Esper et al., 2002). In addition to providing dynamical memory for information extracted from proxy records, the online PDA approach also produces state-dependent statistics, which spread new proxy information in space and to fields other than those directly related to the proxy. This approach allows the dominant modes of climate variability to evolve in time, in contrast to CFR techniques that are derived from one time period (e.g., the instrumental era, or a GCM simulation of the last millennium).

In this work, we present a substantial improvement to previous PDA methods for CFR with the addition of a simple coupled-climate forecast model to propagate assimilated information in time. Including a model provides a more dynamically consistent coupled reanalysis of atmosphere–ocean fields spanning the last millennium, showcasing the breadth of new field information obtainable from climate proxies, and highlighting features of interest in the long-term temperature and upper-ocean heat content behavior. Section 2 describes the background for data assimilation (DA) and its application for CFRs. Section 3 describes the configuration of reconstruction experiments, including the forward model for the proxy measurements, chosen proxy data, and climate simulation data sources. The results of coupled reconstructions are presented in Section 4 with comparisons to previous reconstructions and validation of ocean fields against Instrumental Era products. We present an illustrative investigation of the climate periods known as the Medieval Climate Anomaly (MCA) and Little Ice Age (LIA), and atmosphere–ocean states related to warm periods over Europe in Section 5. Finally, Section 6 compares the temporal characteristics of online reconstructions with offline and simulation output from GCMs.

2 Data Assimilation Background

In this study, we implement an online version of the Last Millennium Reanalysis framework (LMR; Hakim et al., 2016; Tardif et al., 2019), an open-source codebase for CFR. The framework implements variants of the ensemble Kalman filter (EnKF) to optimally blend information from the proxy record with climate model data. The Kalman update equation, defined as

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y} - \mathcal{H}(\mathbf{x}_b)], \tag{1}$$

describes an updated state (e.g., the climate fields of interest), $\mathbf{x}_a$, as a combination of the prior state estimate, $\mathbf{x}_b$, with new information from the innovation, $[\mathbf{y}-\mathcal{H}(\mathbf{x}_b)]$, and the Kalman gain, $\mathbf{K}$. The innovation is the difference between observations ($\mathbf{y}$; e.g., proxy measurements) and estimated observations ($\mathcal{H}(\mathbf{x}_b)$) where $\mathcal{H}$ is a function mapping from state space of the prior into the proxy space. Hereafter, the estimated observations are denoted as $\mathbf{y}_e$ for simplicity. The Kalman gain,

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1}, \tag{2}$$

weights the new contributions from the innovation considering uncertainty in the prior and observations. Here, $\mathbf{B}$ represents the prior covariance matrix, $\mathbf{H}$ is the linearization of $\mathcal{H}$ about the prior ensemble mean, and $\mathbf{R}$ is the proxy error covariance matrix. The Kalman gain term is the key component that translates information from point observations into full fields through estimated covariance structures. In the LMR framework, the general DA algorithm used is as follows:

1. Generate a prior estimate, $\mathbf{x}_b$, for the current time period. In this case, the prior state is an ensemble of climate fields joined along a spatial dimension.
2. Calculate $\mathbf{y}_e$ values using the associated field(s) from the state, $\mathbf{x}_b$, and a proxy system model (PSM). For example, a PSM may use a temperature and/or moisture measure from a grid cell to convert to an estimated tree-ring width.
3. Assimilate available proxy information, $\mathbf{y}$, for the current analysis time using the Kalman update equation (Eq. 1).

2.1 Online Paleoclimate Data Assimilation

In order to assimilate observations over time, we require a method to generate relevant prior state estimates as a "first guess" of the climate state. In the offline case of PDA, the prior estimate is often taken as a random draw of states (centered about the climatological mean) from a long-running climate simulation. The same draw is then used as a prior for each assimilation period, which results in the reconstructed variability over time being entirely determined by the assimilated proxy observations. Online assimilation offers a method to propagate information over time by using short-term forecasts to generate prior estimates for subsequent times. In doing so, some of the memory of previous proxy information is retained through time. However, online forecasting for PDA requires large ensembles of millennium-scale climate simulations, posing an exceptional computational hurdle for most climate model implementations.

To make online PDA for ensemble-based techniques feasible, previous work explored ways to reduce the computational expense while still retaining a skillful model. Reconstruction studies using a particle filter method of ensemble PDA reduce the computational expense by incorporating forecasts from Earth system models of intermediate complexity (e.g., Crespin et al., 2009; Goosse et al., 2010; Goosse, 2017; Dubinkina et al., 2011), or by using a coarsened resolution GCM forecasting at decadal timescales (Matsikaris et al., 2016b). However, research suggests that online data assimilation using a particle filter does not necessarily provide a benefit over the offline method for decadal surface temperature reconstructions (Matsikaris et al., 2015, 2016a). Perkins and Hakim (2017) investigate annual-timescale CFR skill using an ensemble-Kalman-filter-based PDA method with forecasts of the surface temperature from an empirically fit linear inverse model (LIM; Penland & Sardeshmukh, 1995). They find the inclusion of a LIM calibrated on climate model output improves reconstructions of Instrumental Era surface temperatures compared to the offline method, and that it retains computational expediency of the offline method.

A LIM provides an empirically derived encoding of system dynamics into two components: slow-timescale deterministic linear dynamics, and stochastic noise representing nonlinearity and unresolved fast-timescales (Eq. B1). The timescale separation of deterministic and stochastic components is analogous to large-scale climate dynamics forced by weather. As such, LIMs have been widely used to explore mechanisms of atmosphere–ocean interactions such as ENSO and the Pacific Decadal Oscillation (PDO; e.g., Alexander et al., 2008; Newman et al., 2011), and as a forecast skill benchmark for decadal surface temperature (Newman, 2013). For the purposes of PDA, a LIM provides a low-cost mechanism to create a stable model approximating the behavior of sophisticated GCMs and to generate ensembles without requiring any complex state initialization strategy. In a follow on to Perkins and Hakim (2017), Perkins and Hakim (2020, hereafter denoted as PH20) describe the use of a LIM as a GCM analog, providing a general method of calibrating a LIM from a coupled global climate model for ensemble forecasts of multivariate atmosphere–ocean states. They find the multivariate LIM to be skillful out to multi-year lead times and that it reproduces free-running statistics of large-scale climate dynamics indices related to ENSO and the PDO. We use the LIM developed in PH20 as the forecast model for the online PDA reconstructions we present in this study. For a technical description of a LIM, forecasting, and our calibration strategy please see Appendix B.

3 Reconstruction Configuration and Data

In this study, we perform reconstructions covering a period over the last millennium from 1000–2000 C.E. The climate dynamics used to reconstruct the state for each experiment depends on the climate model data used to calibrate the LIM. We select two sources of dynamical information for LIM calibration from the Coupled Model Intercomparison Project phase 5 (CMIP5; Taylor et al., 2012) "Last Millennium Experiments": the Community Climate System Model version 4 (CCSM4; Landrum et al., 2013) and the Max Planck Institute Earth System Model (MPI; Giorgetta et al., 2013). The Last Millennium Simulations cover 850–1850 C.E. and include estimated forcing from greenhouse gases, aerosols (primarily volcanic), solar variability, and land-use changes. The use of two models to reconstruct climate states allows us to assess the robustness of reconstruction results. Our coupled reconstructions include the following fields: 2 m surface air temperature (TAS), precipitation (PR), sea-level pressure (SLP), 500 hPa geopotential height (ZG500), outgoing top-of-atmosphere (TOA) longwave (RLUT), outgoing TOA shortwave (RSUT), sea-surface temperature (SST), sea-surface salinity (SSS), dynamic ocean surface height (ZOS), and 0-700 m ocean heat content (OHC700m).

All fields are regridded to a regular 2° x 2° latitude-longitude grid using bilinear interpolation and are averaged annually from April to March, except for TAS, which has additional seasonal averages. Unlike previous offline reconstruction studies (e.g., Hakim et al., 2016; Tardif et al., 2019), we need to include sub-annual TAS information in the state to calculate the estimated observations ($\mathbf{y}_e$) after each climate forecast. We chose to add all seasonal-average data to the state as individual fields rather than using individual calendar months and forming seasonal averages at runtime. The explicit incorporation of seasonal-average data carries the benefit of decreasing noise in the covariance estimates, which impacts assimilation and forecast skill (e.g., Tardif et al., 2014). All fields are detrended to remove long-term climate model drift and converted to anomalies for LIM calibration and reconstruction procedures.

After preprocessing the data, we use all fields to calibrate the LIM (Eqs. B2, B3). We formulate the LIM and perform forecasts in a reduced parameter space using a two-step empirical orthogonal function (EOF) reduction that PH20 show efficiently preserves shared aspects of large-scale field variability (Appendix B2). Based on LIM skill testing, we retain 40 degrees of freedom for the CCSM4-LIM (20 multivariate and 20 OHC EOFs) and 47 for the MPI-LIM (27 multivariate and 20 OHC EOFs). During the reconstruction process, analysis fields ($\mathbf{x}_a$) are projected into this EOF space, forecast for 1-year using Eqs. B4 and B5, and then back projected into physical space to be used as the next prior state ($\mathbf{x}_b$).

In the LMR framework (e.g., Tardif et al., 2019), the update equation (Eq. 1) involves the use of an ensemble square-root filter approach (Whitaker & Hamill, 2002), which serially updates the state one proxy at time. The serial update implementation is relatively straightforward and allows for covariance localization, which is useful to reduce

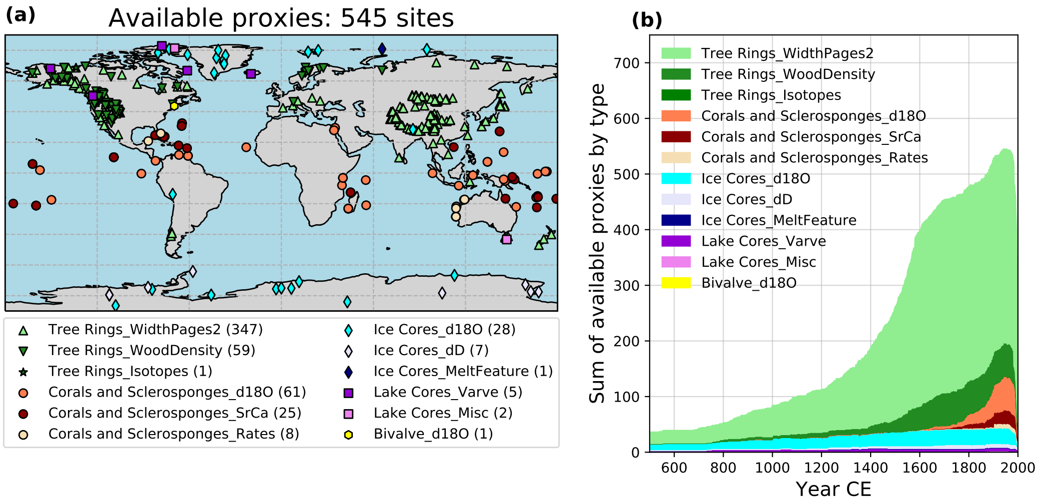

Figure 1. Proxy record spatial distribution (a) and total number of available records separated by type (b) from the PAGES 2k Consortium (2017) database. Proxies shown are those with valid PSMs, which require at least 25-years of overlapping observations with GISTEMP v4 data for calibration.

the effects of sampling error at large distances from observations. In the online case, carrying the fields required for seasonal PSMs increases the state size and computational expense considerably. To speed up the reconstructions, we solve the update equation using a new vector solver variant described in Appendix A. This technique reduces the update problem to the smallest possible space spanned by the ensemble size and number of proxies, at the expense of no covariance localization on the full-space state (a commonly used measure to control spurious long-range correlations).

The last important factor concerning experiment configuration is the selection of proxies and proxy system models (PSMs; represented as $\mathcal{H}$) used to predict the proxy values from the climate state variables. We use proxy records with annual time resolution from the PAGES 2k Consortium (2017, PAGES2017) database with a recently updated Palmyra coral record (Emile-Geay et al., 2013b; Anderson et al., 2019). The PAGES2017 database is a quality-controlled compilation of metadata and proxy records screened for temperature sensitivity. For each proxy PSM, we fit a linear univariate model using objectively determined seasonal averages (as in Tardif et al. (2019)) for tree-based proxies and expert-derived seasonality (PAGES 2k Consortium, 2017) for all other proxy types. (See Appendix C for a description of the objective testing procedure.) The PSMs are fit against co-located instrumental temperature data from the NASA Goddard Institute for Space Studies Surface Temperature Analysis GISTEMP v4 (Hansen et al., 2010) dataset (similarly regridded to a 2° x 2° grid). Only proxies with an overlap of at least 25-years with instrumental data are calibrated, which results in 545 usable proxies (Fig. 1).

With the calibrated LIM and PSMs, we perform Monte-Carlo (MC) iteration reconstruction experiments where in each case a 100-member ensemble and 75% of the available proxy data are assimilated over all times. We run 50 realizations using this resampling strategy to assess uncertainty related to proxies, such as changing coverage and potential dating errors, and LIM climate forecasts. Previous work finds that iterative reconstructions in this manner provide beneficial results (Tardif et al., 2019). The use of 100 ensemble members is consistent with previous work that shows ensembles of this size reasonably sample field covariances used in the Kalman gain (Hakim et al., 2016; Tardif

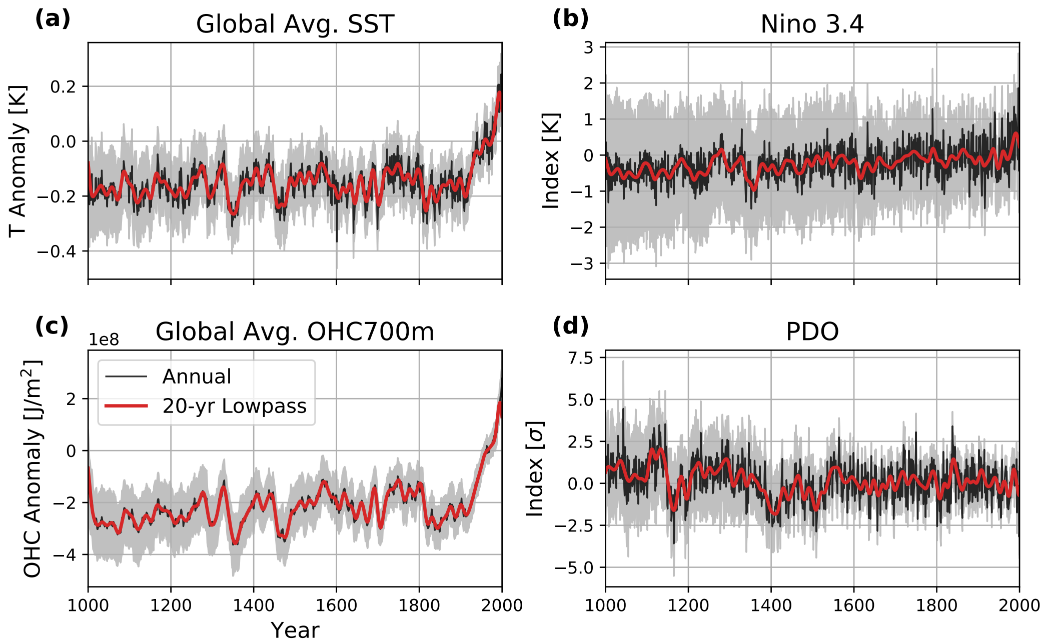

Figure 2. A selection of scalar index ensembles calculated from the annual-average field reconstructions using the CCSM4-LIM including: global average (a) sea-surface temperatures (SST) and (c) 0–700 m ocean heat content (OHC700m), and dynamic indices of the (b) Niño 3.4 region average SST and (d) Pacific Decadal Oscillation (PDO) index. Curves represent the annual ensemble-mean indices taken over all 50 × 100 members (black) and the associated 95% confidence region (grey shaded), and the smoothed 20-year lowpass filter of the ensemble mean (red).

et al., 2019). In addition to online forecasts using LIMs calibrated on the CCSM4 and MPI Last Millennium Simulations (hereafter referred to as CCSM4-LIM and MPI-LIM), we perform an offline reconstruction for comparison, which uses climatological draws from the CCSM4 simulation as a prior. We use a seeding strategy to ensure that the proxy samples and initial prior ensemble samples are equivalent for a given MC iteration between the CCSM4-LIM, MPI-LIM and offline experiments.

4 Last Millennium Reconstruction using Online Data Assimilation

As an overview of reconstructed ocean results, we show area-weighted global average temperature and OHC700m, the Niño 3.4 index, and the Pacific Decadal Oscillation (PDO) index (Fig. 2). The Niño 3.4 index is the average SST over the region from 5S–5N and 170W–120W, and the PDO is calculated by projecting the first EOF of North Pacific (20N–70N and 110E–110W) detrended SST variability from the CCSM4 last millennium simulation onto the reconstructed SST field. We also provide ensemble-mean scalars that are smoothed using a 61-sample Lanczos filter (Duchon, 1979) with a cutoff frequency of 20-years to highlight low-frequency variability in the data. All reconstructed data are centered with a reference period of 1951–1980 unless stated otherwise.

Reconstructed global average SST (Fig. 2a) portray relatively cool conditions for the majority of the last millennium with apparent decadal-to-centennial-scale (hereafter referred to as dec-cen) variability (e.g., temperature swings from 1300–1500 and repeated volcanic cooling events post 1600 C.E). The average SST warms considerably during the

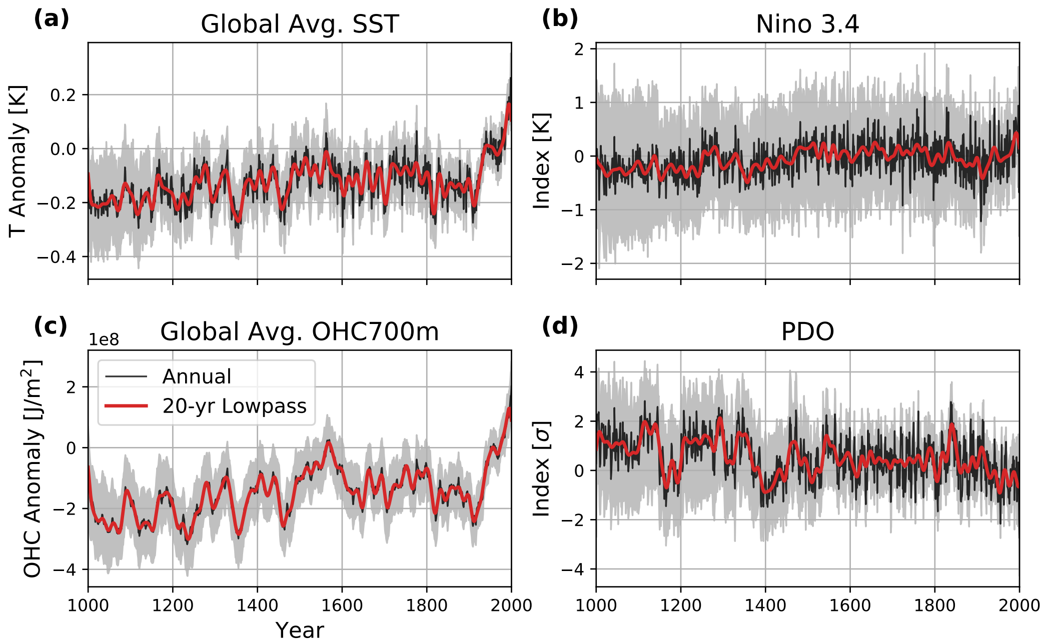

Figure 3. As in Fig. 2 but for the MPI-LIM reconstruction.

Instrumental Era ending ∼0.4 K warmer than the 1000–1850 CE average of approximately −0.2 K. Additionally, the ensemble-mean SST appears to exit the pre-industrial reconstructed range of millennial temperatures by the first half of the 20th Century. Notably absent are global-scale warm anomalies in the early period typically associated with the Medieval Climate Anomaly (MCA) or millennial-scale cooling present in other field reconstructions (e.g., Mann et al., 2009; Hakim et al., 2016). We investigate the expression of the MCA in our reconstruction in Section 5. The reconstructed upper-ocean heat content (Fig. 2c) shows less interannual variability than SST, but similar dec–cen variability, which corresponds to the much higher thermal inertia of an ocean layer. Concerning OHC700m thermal inertia, the first few decades of OHC700m show evidence that there is some "spin-up" time associated with online DA for this field, as the state is drawn toward observations.

The dynamic indices of ENSO (Niño 3.4) and the PDO display the reconstructed character of known predominant modes of natural variability. The Niño 3.4 index (Fig. 2b) shows relatively stable temperatures over the last millennium with the ensemble average, suggesting a slight warming trend over time (∼0.5 K / 1000 yrs). Also evident in the Niño 3.4 reconstruction is an increase in the positive temperature trend during the modern era, consistent with global warming. The large span of the 95% confidence region early in the period is because the Niño 3.4 index is a regional-scale index and there are relatively few local constraints before the 1600s when more coral records become available. The PDO index (Fig. 2d) also suggests some dependence on available proxy observations. Before 1600 C.E., there is pronounced dec-cen variability in the ensemble-average PDO index with fluctuations of about 2–3 σ during some periods. After the year 1600, a large number of tree-based information and some corals (Fig. 1) become available and the PDO index shows reduced low-frequency variability (Fig. S1). Only considering reconstructed PDO values after 1600 C.E., there is no distinct long-term trend or changes to variability between the pre-industrial and modern period.

Scalar indices calculated from the MPI-LIM reconstruction (Fig. 3) give qualitatively similar results as the CCSM4-LIM experiment. Lowpass- filtered scalar correlations with the CCSM4-LIM reconstructions are 0.91 for global average SST, 0.85 for OHC, 0.83 for the PDO, and 0.68 for the Niño 3.4 index, respectively. The temperature-related indices (SST, OHC700m, and Niño 3.4) show some warming into the 16th century followed by cooling into the 17th and 18th century. The warming of the MPI-LIM ensemble-averages (SST and OHC700m) into the 1500s reaches similar temperatures as during the early 20th-Century although the warming occurs over a longer period. Additionally, the OHC700m warm-period before 1600 C.E. is outside of the CCSM4-LIM confidence intervals. This may suggest some underestimation of the LIM forecast variance. The PDO shows large fluctuations in the average index value prior to the 1600s, with similar phasing as in the CCSM4-LIM case. Similarity in the long-term PDO index fluctuations suggest the two online reconstructions reproduce the same ocean state given the same proxy data. However, the change in dec–cen variability after 1600 C.E. is likely due to the sparsity of the proxy data before that time. To illustrate the character of spatial variability from the ocean–atmosphere fields, we provide videos of the evolution of selected fields and the geographic distribution of available proxies over time (Movies S1 and S2). In these field sequences, the coupled variability associated with ENSO and with slower modes of variability in the North Atlantic and Pacific are prominent features, as well as the global-scale warming during the 20th Century.

4.1 Instrumental Validation

We now validate the field reconstructions during the instrumental period for SST and OHC700m spatial fields and related scalar quantities. Comparison products include the following: Hadley EN4 v4.2.1 (Good et al., 2013), a quality-controlled and objectively interpolated dataset based on ocean profile measurements from 1940–2000 C.E., the Geophysical Fluid Dynamics Laboratory Ensemble Coupled Data Assimilation (GFDLECDA; Chang et al., 2013), a coupled climate model reanalysis using ocean observations from 1960–2010 C.E., and the European Center for Medium-range Weather Forecasting (ECMWF) ORA-20C dataset covering 1900–2010 (de Boisséson et al., 2018), an ocean-field reanalysis for use as initial conditions in ECMWF coupled reanalysis product (Laloyaux et al., 2018). For El Niño, PDO, and OHC comparison, we also use the Earth System Research Laboratory Niño 3.4 time series (1950–2018, accessed Apr. 19, 2019), the Mantua et al. (1997) PDO index hosted by the Joint Institute for Study of the Atmosphere and Ocean (research.jisao.washington.edu/pdo/PDO.latest.txt, accessed Apr. 19, 2019), a gridded OHC estimate optimally interpolated using information from CMIP5 historical simulations, and an estimate of OHC over the full Instrumental Era using observed SSTs and a passive ocean transport model (Cheng et al., 2017; Zanna et al., 2019).

Temporal gridpoint correlations with instrumental products show large-scale agreement with SST and more regionally dependent agreement for OHC700m (Fig. 4). SST correlations (Fig. 4, column a) are largely positive in the tropics, especially the tropical Pacific Ocean, less correlated in the Southern Ocean regions and Northwest Pacific, and uncorrelated-to-anti-correlated in the Labrador Sea and North Atlantic region south of Greenland and Iceland. SST correlations display better spatial agreement with the two model-reanalysis experiments (GFDLECDA, ORA-20C) than the observation-only data (HadleyEN4). Correlations for OHC700m (Fig. 4, column b) similarly show the tropical Pacific Ocean as a region with more agreement between the reconstruction and instrumental products. Additionally, OHC700m correlations are moderately positive in the subtropical Atlantic Ocean and weak-to-moderately positive in the mid-latitude Southern Ocean areas when comparing with the four instrumental products. The HandleyEN4 dataset shows the lowest correlations, especially in the Southern Hemisphere, with the CCSM4-LIM reconstructed OHC700m. In general, more regions of small or negative correlations are apparent for OHC700m, but the North Atlantic and Labrador Sea region again is a common region of anti-correlation with instrumental products. We note that

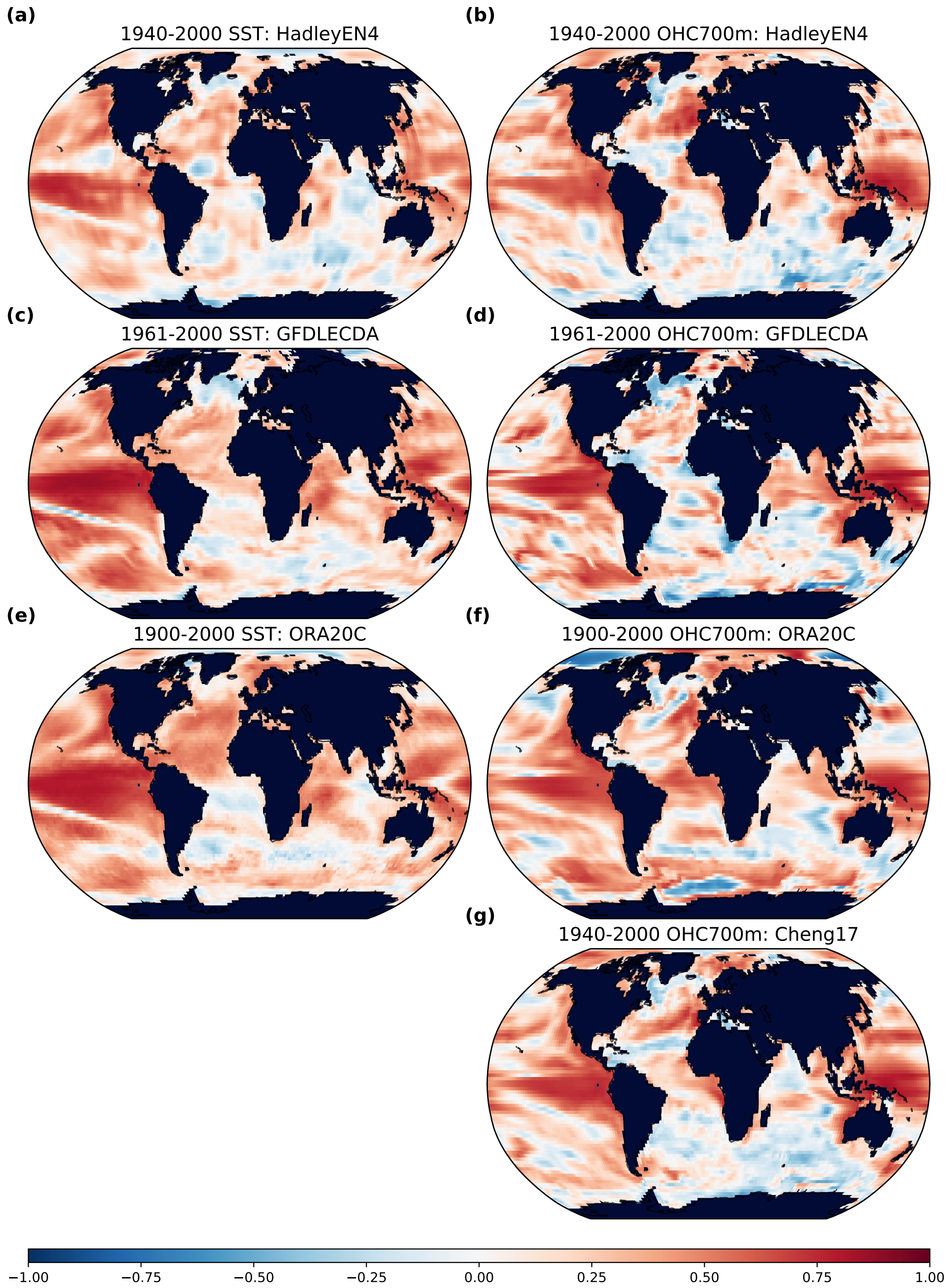

Figure 4. Detrended temporal gridpoint correlations of the LMR Online (CCSM4-LIM) reconstruction with Instrumental Era observational and reanalysis products for SST (column a) and OHC700m (column b). Correlations are calculated against Hadley EN4 data (a, b; 1940–2000), GFDLECDA (c, d; 1961–2000), ORA-20C (e, f; 1900–2000), and Cheng2017 (g; OHC only, 1940–2000)

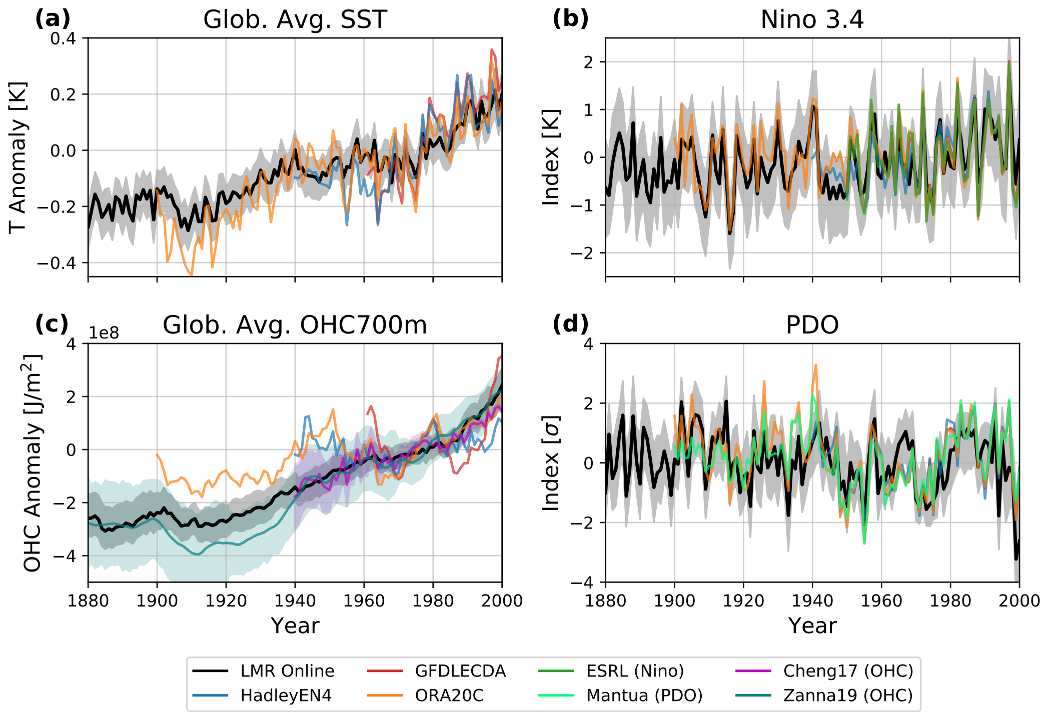

Figure 5. Scalar index comparison between the LMR Online (CCSM4-LIM) reconstruction (black with 95% confidence bounds in grey shading) and instrumental products for (a) SST, (b) Niño 3.4, (c) OHC700m, and (d) PDO. The HadleyEN4, GFDLECDA, and ORA-20C products are compared in all cases. Additionally, ESRL Niño 3.4 data, the Mantua et al. (1997) PDO index, and Cheng et al. (2017) and Zanna et al. (2019) OHC data are compared. Error bounds ($\pm 2\sigma$) are shown for the Cheng17 and Zanna19 OHC700m data.

OHC observations were generally sparse during the 20th century before the implementation of the ARGO observing array during the late 1990s (Riser et al., 2016). Therefore, observationally-based spatial products have large uncertainties for OHC700m during the 20th Century and considerable discrepancies (e.g., see Fig. S2 for instrumental product spatial correlation comparisons). The southern Atlantic and Indian oceans show up as a region of low correlation for both SST and OHC700 likely due to the large distance from the assimilated proxy observations. However, this is also the region that displays the most uncertainty between observational products (e.g., Fig S2). We also validate the MPI-LIM SST and OHC700m against these products and find similar correlation patterns as those described for CCSM4-LIM reconstructions (Fig. S3).

Figure 5 shows a comparison of instrumental scalar indices, and Table 1 shows the associated correlation values. The reconstructed global average SST values (Fig. 5a) follow the decadal trajectory of the instrumental products but also show smaller amplitude interannual variability. Additionally, the small 95% confidence interval for SSTs relative to the validation data mismatch suggests the ensemble variance of this scalar measure is underestimated. SST correlations with the instrumental products are generally high with values between 0.8–0.9. As in the spatial comparison, the LMR Online reconstruction agrees best with the two reanalysis products (GFDLECDA and ORA-20C). Despite the lack of interannual anomaly amplitude for global-average SST, the Niño 3.4 index (Fig. 5b) matches the comparison products well for both phase (correlations between 0.8–0.87) and amplitude. Upper-ocean heat content products (Fig. 5c) have more dis-

Table 1. LMR Online (CCSM4-LIM) reconstruction scalar correlations with instrumental products.

Product	Glob. Avg. SST	Glob. Avg. OHC700m	Nino 3.4	PDO
HadleyEN4	0.80	0.27	0.79	0.61
GFDLECDA	0.89	0.58	0.87	0.56
ORA20C	0.90	0.79	0.80	0.63
ESRL	–	–	0.84	–
Mantua	–	–	–	0.58
Cheng17	–	0.94	–	–
Zanna19	–	0.98	–	–

agreement among them, but the LMR Online reconstruction is strikingly similar to the Cheng17 (correlation of 0.94) and Zanna19 (correlation of 0.98) data. These two data products share the distinction of using a passive style of model–observations blending. The other products (HadleyEN4, GFDLECDA, and ORA-20C) tend to have much larger decadal-scale variability and differences of the global average OHC700m trajectory. Finally, the PDO comparison (Fig. 5d) shows that the CCSM4-LIM reconstruction reproduces the decadal-scale changes of the PDO, but is less skillful for interannual PDO variability (correlations near 0.6 for all products). The 95% confidence interval is generally close to encompassing the interannual instrumental PDO data, which suggests the uncertainties of the reconstructed PDO estimate are reasonable.

The MPI-LIM reconstruction scalar validation (Fig. S4) shows similar global average performance based on correlation but less skill for the dynamic indices. The reconstructed global average SST and OHC700m show a noticeable pause in warming from 1940–1970 C.E. and do not warm as strongly towards the end of the reconstruction period. The reconstructed Niño 3.4 in the MPI-LIM case has smaller amplitude interannual anomalies, and correlations with instrumental products of 0.72–0.78 (Table S1). The reconstructed PDO, in this case, shows less clear agreement with decadal-scale variability and the correlations for the shorter-length comparisons (HadleyEN4, GFDLECDA, Mantua) decrease to 0.24–0.42. The differences in reconstruction performance of dynamic indices suggest the MPI-LIM produces less representative reconstructed fields in these regions, which could be related to the character of the MPI-LIM dynamics, or how proxy information is weighted given the forecast ensemble characteristics.

4.2 Comparison to previous reconstructions

To put our reconstruction in the context of previous research, we show the LMR Online (CCSM4-LIM) reconstruction compared to other proxy-based reconstructions (Fig. 6). We apply a 20-year lowpass filter to all data except for SST, which is averaged to 100-year intervals to correspond with the SST estimates from McGregor et al. (2015). Note that the comparison SST data (McGregor et al., 2015) are re-averaged to 100-year intervals (as opposed to the 200-year intervals they present) using that study's associated code and data.

The Northern Hemisphere (NH) average temperature reconstructions (Fig. 6a) show closer correspondence from roughly 1500 C.E. onwards with LMR Online on the cooler end of the distribution. Notable cooling events in the NH temperature (e.g., in the 1300s and 1400s) display a more extensive range of hemisphere average temperature variability and the LMR temperatures are generally 0.25–0.5 K cooler than other reconstructions in the early portion of the last millennium. Cooler NH temperatures suggested by our reconstruction are closer in agreement with borehole estimates (Pollack & Smerdon,

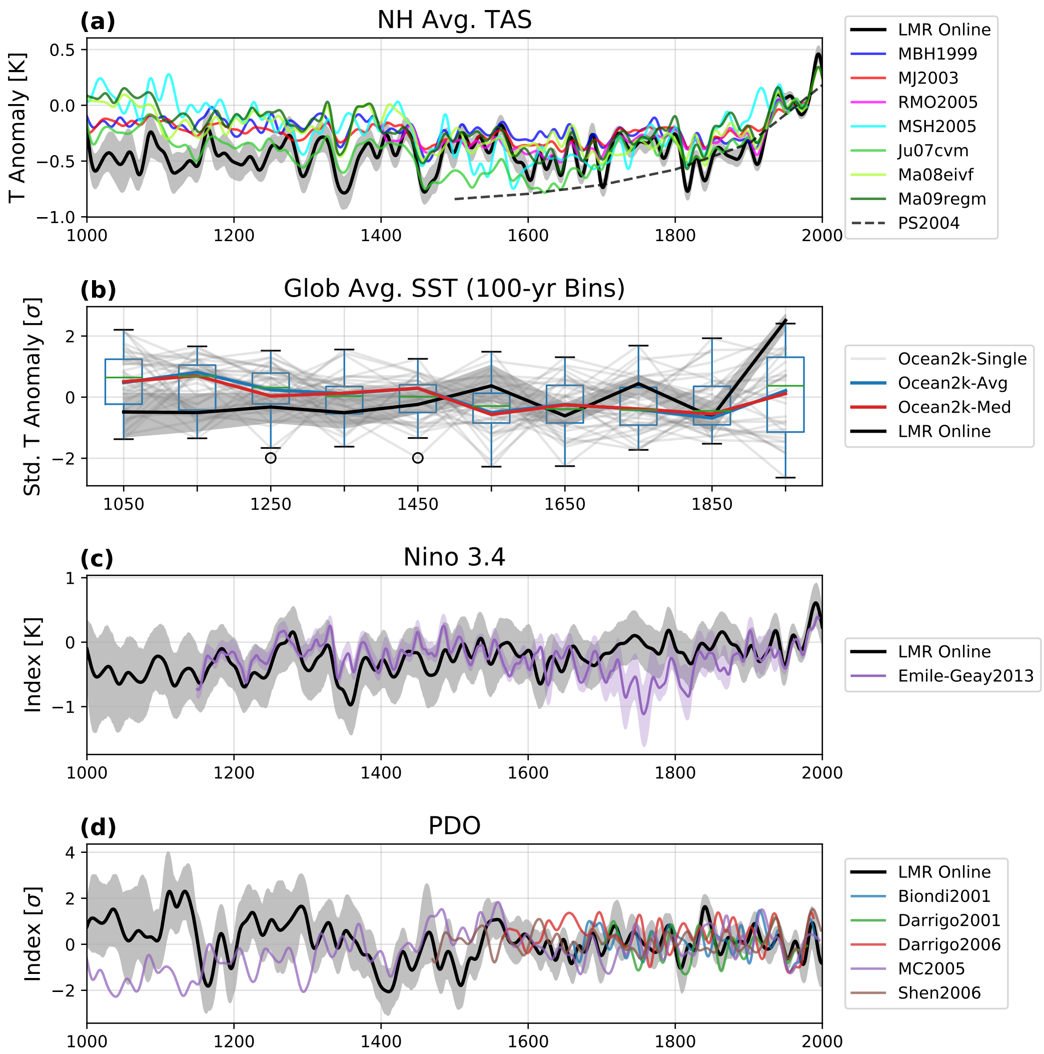

Figure 6. A comparison of LMR Online (CCSM4-LIM) reconstructed scalar indices with previous reconstructions. Northern hemiphere average TAS (a) is compared between LMR Online (black; grey shading for 95% confidence region) and the following studies: MBH1999: Mann et al. (1999), MJ2003: Mann and Jones (2003), PS2004: Pollack and Smerdon (2004), RMO2005: Rutherford et al. (2005), MSH2005: Moberg et al. (2005), Ju077cvm: Juckes et al. (2007), Ma08eivf: Mann et al. (2008), and Ma09regm: Mann et al. (2009). Global average SST (b) is compared with McGregor et al. (2015), a compilation of 57 100-year bin-averaged sediment records with the basin-weighted mean standard anomaly (blue), median standard anomaly (red), and box-whisker plots displaying the inner-quartile range (IQR) with 1.5×IQR whiskers at each time interval. The Niño 3.4 index (c) is compared against index reconstructions from Emile-Geay et al. (2013a) where the solid line is the average of three reconstructions they present and shading denotes the range. The PDO (d) is compared with collection of index reconstructions using proxies with a variety of regional coverage (Biondi et al., 2001; D'Arrigo et al., 2001; MacDonald & Case, 2005; D'Arrigo & Wilson, 2006; Shen et al., 2006).

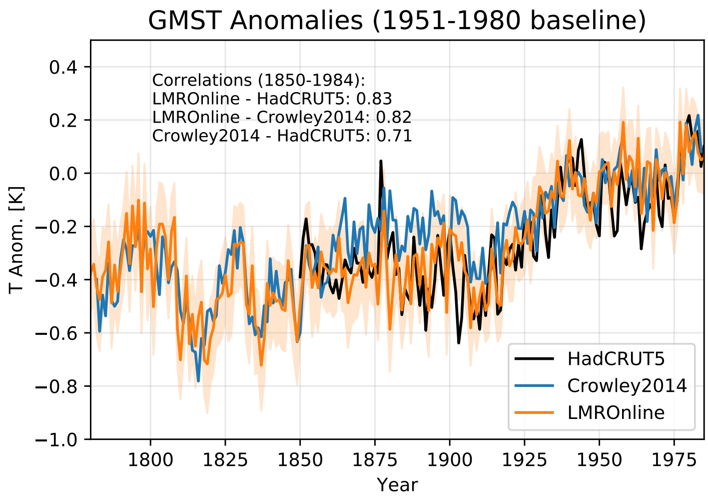

Figure 7. A comparison of LMR Online (CCSM4-LIM) reconstructed global mean surface air temperature (orange with 95% confidence region shaded) against the Crowley et al. (2014) ENSO-tuned global composite reconstruction (blue) and HadCRUT5 (Osborn et al., 2021) observational data (black).

2004), and the tree-based estimates from Juckes et al. (2007). Global average SST (Fig. 6b) estimated from sediment cores generally has a wider uncertainty range than the LMR Online reconstructed SST owing to fewer single-location measurements and large differences in temporal recording resolution of the different cores. Throughout the 1000-year reconstruction, our reconstructed SST values fall within the whiskers (1.5$\times$ inner-quartile range) for all times except during the last 100-year interval where global warming has considerable influence, and sediment cores lack information. LMR Online global average SSTs are cooler earlier in the reconstruction period compared to the Ocean2k median and mean values.

The Niño 3.4 index shows limited agreement with Emile-Geay et al. (2013a, referred to as EG13) while the PDO index shows general disagreement with the five comparison studies (Biondi et al., 2001; D'Arrigo et al., 2001; D'Arrigo & Wilson, 2006; MacDonald & Case, 2005; Shen et al., 2006) outside of the Instrumental Era. Comparing the Niño 3.4 index (Fig. 6c), the EG13 reconstruction is mostly within the uncertainty bounds of the LMR Online reconstruction except during the period from approximately 1700–1900 where the EG13 reconstructed temperatures are about 0.5–1.0 K cooler. The period of divergence (1700–1900) occurs over a time where the number of available proxy records markedly increases in both reconstructions. We note that the EG13 large-scale tropical multiproxy network does overlap with the records used in our reconstruction experiments, but the number of used proxies in the EG13 network is much smaller (36 total records). The PDO index comparison (Fig. 6d) shows that the reconstructions mostly disagree in signal phasing. Correlations between the experiments (Fig. S5) show the D'Arrigo et al. (2001) reconstruction has the highest agreement with the CCSM4-LIM reconstructed PDO with a value of $\sim$0.4. Otherwise, most correlations are quite low or even anti-correlated (e.g., MacDonald & Case, 2005). The Biondi et al. (2001) reconstruction has the broadest agreement across records, but even that agreement is relatively low (correlations between 0.2–0.5).

For additional perspective, we include a comparison of annual global-mean surface temperature from our online DA reconstruction with that of Crowley et al. (2014), which they describe as a "baseline" reconstruction employing minimal proxy processing. The Crowley et al. (2014) method uses a geographically broad and fixed proxy network over time. Proxy data are centered, standardized, zonally averaged, and then linearly regressed against instrumental temperature data (HadCRUT3) to reconstruct temperature from the proxy composites before the instrumental period. This definition is analogous to the inverse of the $\mathcal{H}$ operator in (Eq. 1), taking climate as a linear function of the proxies ($\mathcal{H}$ relates proxy values linearly to climate). Overall, the two reconstructions are quite similar with a correlation of 0.82 over 1850–1984 C.E., in-phase multi-decadal variability, and a similar magnitude of overall warming into the late 20th century. The two reconstructions slightly diverge around 1860–1910 C.E. where the Crowley reconstruction is about 0.1–0.2 C warmer. When comparing to HadCRUT5 instrumental temperature data (Osborn et al., 2021), the LMR Online reconstruction outperforms the Crowley baseline with a higher correlation (0.83 compared to 0.71 for Crowley) and better agreement with observations through 1860–1910 C.E. where Crowley data shows slightly higher temperatures. We emphasize that the Crowley composite reconstruction is calibrated directly to large-scale zonal averages of HadCRUT3 data, while our reconstruction derives indices from reconstructed full-fields informed by locally assimilated proxy data.

5 Medieval Climate Anomaly in LMR Online

The "Medieval Climate Anomaly" (MCA) is an often targeted period (e.g., 950–1250 C.E.) for investigating the mechanisms and magnitude of natural climate variability before the Industrial Era (see review by Diaz et al., 2011). Documentary evidence (Lamb, 1965) and reconstructions based on proxy records (e.g., Mann et al., 2008, 2009; Ljungqvist, 2010) suggest the possibility of extended regional-to-hemispheric warm periods during the MCA. Mann et al. (2009), using a multi-proxy statistical technique, estimate the spatial character of the temperature transition from the MCA into the "Little Ice Age" (LIA), a period of cool climate conditions roughly between the 1400s and 1800s. They find the MCA–LIA difference is defined by broad warmth with a La Niña-like temperature pattern in the tropical Pacific, and that potential mechanisms for this pattern are related to forcing from ENSO and high-latitude atmospheric circulation variability. Goosse, Crespin, et al. (2012) and Goosse, Guiot, et al. (2012) examine climate dynamics of the MCA period over Europe and the northern hemisphere using the ensemble particle filter PDA method. Spatial reconstructions coupled with information from regional proxy assessments suggest possible mechanisms for the MCA–LIA transition including changes in North Atlantic SST and atmospheric circulation (e.g., Trouet et al., 2009), teleconnections related to cool tropical Pacific temperatures (Cobb et al., 2003; Mann et al., 2009), feedbacks related to solar variability (Ammann et al., 2007; Meehl et al., 2009; Goosse, Crespin, et al., 2012) and volcanic activity (Atwood et al., 2016). However, uncertainties related to the proxy network and reconstruction methodology can produce significant differences in estimated spatial characteristics of the MCA (Wang et al., 2015). Furthermore, recent work utilizing statistical and PDA-based CFR techniques finds little evidence of globally coherent warm or cold extremes before 20th Century warming (Neukom, Steiger, et al., 2019). As an example, we provide a short investigation of the LMR Online reconstruction of the MCA–LIA transition and field relationships with average temperatures over the European region.

The reconstructed NH average temperature, as previously described, shows relatively cool temperatures throughout the pre-industrial period (Figs. 8a, 8b). However, when averaging temperature over the European (EU) region (40–80N, 20W–40E), decadal-scale warming events within the MCA period are apparent. In the CCSM4-LIM reconstruction (Fig. 8a), the EU warm events reach a magnitude about half as large as the early 20th Century warming. The MPI-LIM reconstruction (Fig. 8b) also shows decadal-scale warm periods during the MCA consistent with CCSM4-LIM results.

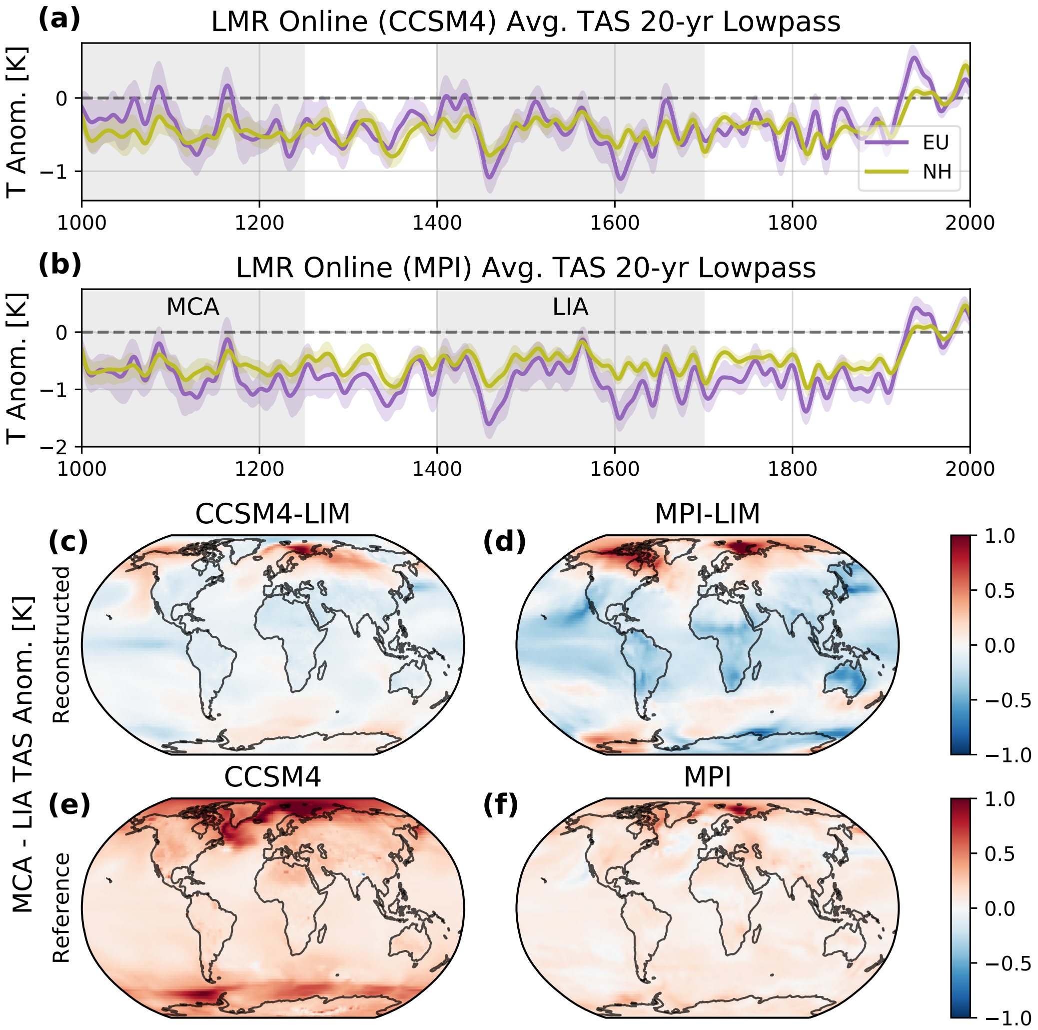

Figure 8. Comparisons between Europe (40–80N, 20W–40E) and North Hemisphere lowpass-filtered average TAS anomalies (centered about 1951–1980 C.E.) for the (a) CCSM4-LIM and (b) MPI-LIM reconstructions. As in Mann et al. (2009), spatial differences between the reconstructed (1000–1250 C.E.) and the LIA (1400–1700 C.E.) time periods are shown for (c) CCSM4-LIM and (d) MPI-LIM reconstructions and for reference (e) CCSM4 and (f) MPI Last Millennium Simulations.

Spatial differences in the reconstructed TAS field between the MCA and LIA periods for the CCSM4-LIM and MPI-LIM reconstruction experiments are shown in Figures 8c and 8d. Both reconstructions have positive temperature anomalies over northern Europe with a maximum occurring over the Barents Sea in the sea-ice transition region, some expression of higher-latitude warming in North America, and colder tropical regions. However, there is no evident global-scale warmth, which agrees with results presented by Neukom, Steiger, et al. (2019) but stands in contrast to the reference GCM simulations (Fig. 8e and 8f). The patterns of the reference simulations for the MCA–LIA difference presents as mostly global-scale warming, whereas the reconstructions are more regionally distinct, with warmer conditions at high-latitudes and colder in the tropics.

Considering the two reconstructed MCA–LIA patterns (Figs. 8c, 8d), there are substantial differences in the spatial character of anomalies outside of Europe. The CCSM4-LIM reconstruction MCA–LIA difference has warming localized to Alaska and Canada with a positive PDO expression in the North Pacific and a La Niña-like state in the tropical Pacific. The MPI-LIM/MCA–LIA difference has broad warming across the entire Arctic region with broadly cold tropical and North Pacific temperatures. Consistent with the sentiment expressed in Wang et al. (2015), differences in the regional expression of temperature anomalies reinforce the notion that result robustness to methodological decisions should be considered along with reconstruction results.

Both the CCSM4-LIM and MPI-LIM reconstructions point to a warm anomaly in the vicinity of northern Europe during the MCA. We now investigate how temperatures in Europe (20-year lowpass filtered) covary with the coupled fields of SST/ZG500 and OHC700m/SLP using regression analysis during 1000–1850 C.E. (Fig. 9). In the CCSM4-LIM reconstruction, warm temperatures in the EU region are related to generally warm SSTs (Fig. 9a) with maximum warmth in the Norwegian and Barents Sea and warmer SSTs in the Gulf Stream region. Upper-level ZG500 field shows an annular pattern of increased mid-latitude heights with maxima situated over the North Atlantic, Europe, North Pacific, and North America. Lower values of ZG500 over the Arctic suggest a strengthening of the upper-level circulation, which is reminiscent of a positive-phase Arctic Oscillation (AO; Wallace & Thompson, 1998). The positive-phase AO is highly correlated with a positive North Atlantic Oscillation (NAO) phase (e.g., Ambaum et al., 2001), which is associated with warmer temperatures and enhanced storminess over the EU region in modern times (Rogers, 1997; Trigo et al., 2002). The regression on the OHC700m field (Fig. 9b) shows similarly located regions of warm anomalies associated with EU-region warmth, and the SLP regression suggests the Arctic circulation anomaly strengthens near the surface. Significance tests show that the SST and OH700m anomalies are significant at the 95% thresholds (Figs. S8a, S8b). Significant aspects of the ZG500 field are associated with the positive height anomaly maxima, but SLP significance is limited to the North Atlantic and the central Arctic regions.

The MPI-LIM reconstruction regressions (Figs. 9c, 9d) mostly disagree with the CCSM4-LIM results over the North Pacific, but display some similar results as in the CCSM4-LIM case in other regions. Positive SST and OHC700m anomalies in the Norwegian and Barents Sea are associated with warm EU average temperature. The OHC700m field has negative anomalies south of Greenland as opposed to more neutral heat content in the CCSM4-LIM reconstruction, but these regions are not significantly related to EU average temperature (Fig. S8c). For ZG500 (Fig. 9c), there are similarly four centers of height local maxima in the mid-latitudes, but in contrast to the CCSM4-LIM case, there is an increase in ZG500 over the Arctic. At lower levels, the SLP regression implies lower pressure over the Arctic region with the local maxima located over Greenland and Arctic regions north of EU and Asia. Lower Arctic SLP and an increase in Arctic ZG500 suggests warmer temperatures in Europe are associated with higher atmospheric thickness over the Arctic. However, the SLP relationships with EU temperature are not

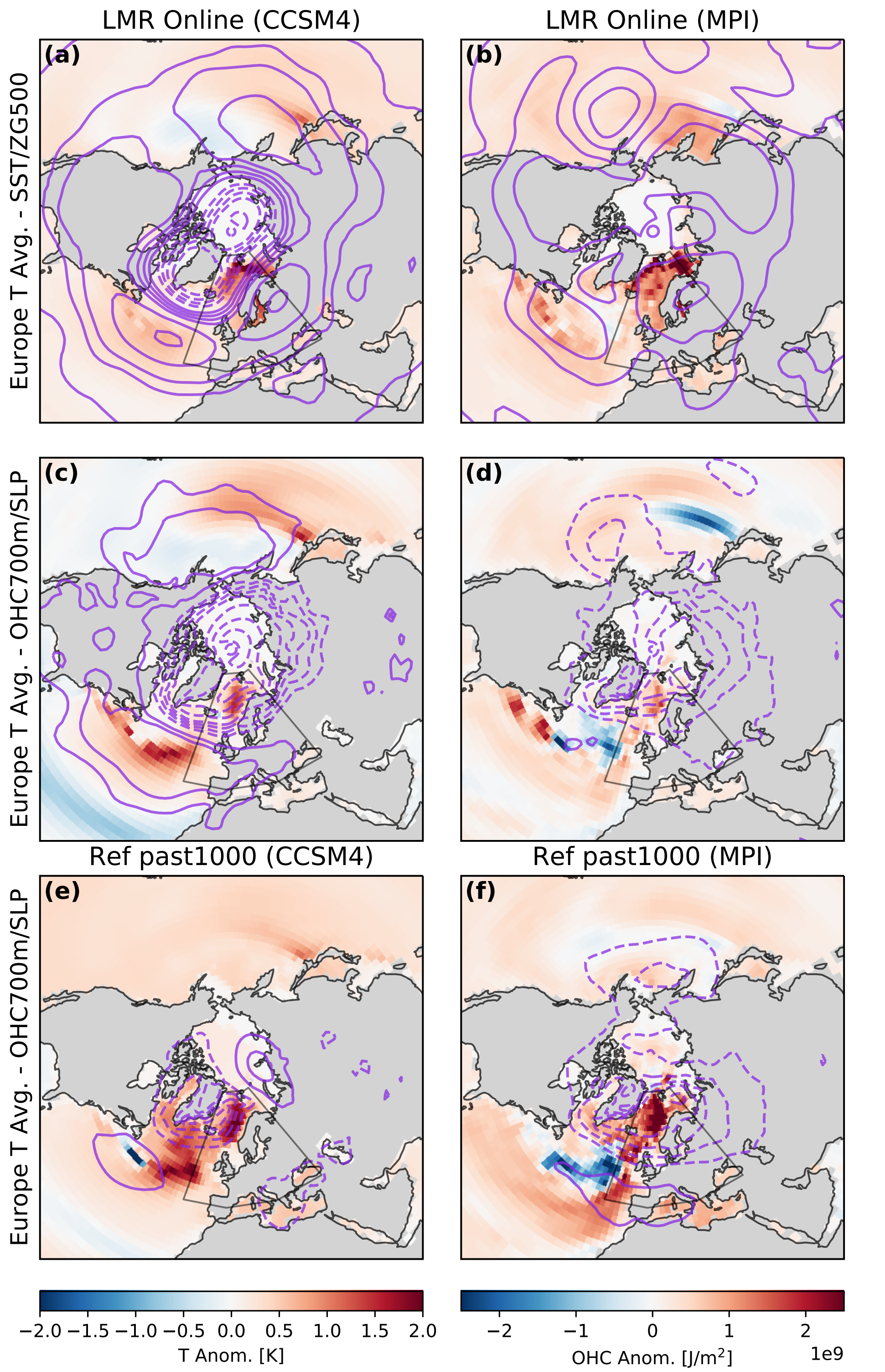

Figure 9. Regression of the Europe average TAS (region denoted by black box) from 1000–1850 C.E. onto fields of (row a) SST/ZG500 and (rows c and e) OHC700m/SLP for the (a, c) CCSM4-LIM and (b, d) MPI-LIM reconstructions, and (e) CCSM4 and (f) MPI reference last millennium simulations. ZG500 field contour levels are incremented by 2.5 m from 2.5–10 m and 5 m from 10–20 m for positive (solid) and negative (dashed) values. SLP field contour levels are incremented by 0.25 hPA from 0.25–1.0 hPa and by 0.5 hPa from 1.0–2.5 hPa. All data are 20-year lowpass filtered prior to calculating the regression.

significant at the 95% confidence level (Fig. S8d). The CCSM4-LIM and MPI-LIM reconstruction regressions suggest a strengthened lower-level Arctic circulation anomaly in the lower atmosphere along with the warmer regional SSTs and OHC700m anomalies. Furthermore, the regression relationships are notably different from those derived from the reference CCSM4 Last Millennium simulations, which tend to be much more regional in scale (Fig. 9e). The change in character of the reconstructed field relationships with EU temperature suggests that proxy assimilation adds new information to coupled-field variability over the past 1000 years.

6 Temporal Constraints in the Online Technique

A motivating factor in pursuing online assimilation for climate reconstruction is the long-term predictability related to ocean dynamics and memory (Goosse, 2017). With predictive skill on timescales longer than the temporal resolution of the proxies (e.g., Hawkins & Sutton, 2009; Perkins & Hakim, 2020), adding a forecast model preserves information assimilated at previous times. In this section, we provide an assessment of changes to reconstructed state memory and dynamics for the ocean fields by comparing the online results with the offline results and the reference GCM simulations. For the offline reconstruction, proxy data is the only information governing variability and memory over time. In contrast, the online reconstructions have LIM dynamics, which persist and constrain climate signals when less proxy data are available. The reference models provide a comparison between the variability that the LIM dynamics are calibrated to emulate, and the influence of the LIM coupled to proxy assimilation. Where applicable, we compare time periods pre- and post-1600 C.E. in order to assess the dynamics with fewer and more proxies, respectively, available to constrain the reconstruction.

Comparing the online CCSM4-LIM results with the offline CCSM4 reconstruction of lowpass-filtered global average TAS, SST, and OHC700m (Fig. 10) reveals cooler conditions ($\sim$1000–1600 C.E.), a tightened confidence interval, and larger dec–cen fluctuations in the online results. Because both the online and offline LMR reconstructions use the same proxy information, TAS and SST signal phasing between the two reconstructions is high (correlations near 0.9). Additionally, the online reconstruction results for TAS and SST generally fall within the confidence interval of the offline results. However, for OHC700m, differences in dec–cen variability are more substantial. Overall, the addition of temporal memory and dynamical information results in cooler temperatures early in the reconstruction period, diminishing the millennial-scale cooling trend compared to the offline case. Even though low-frequency variability may not be well represented by annually-resolved proxies, we recover an estimate of low-frequency variations via assimilation and the slow-timescale dynamics of the LIM.

To test if the cooler reconstructed temperature during the early period of the last millennium are an artifact of the LIM, we performed two additional experiments. First, to test whether the inclusion of volcanic events in the LIM calibration may result in sustained artificially cool temperatures during low proxy-information periods, we performed an equivalent reconstruction experiment using a LIM calibrated on the CCSM4 pre-industrial control (piControl) simulation without forcing (see Section S2 for details). Results show that the piControl-LIM experiment still displays a relatively cool early period with no millennial-scale cooling trend (Fig. S6). Second, we tested sensitivity to the initialization time by starting another CCSM4-LIM reconstruction from the year 1 C.E. This experiment produced nearly identical reconstruction results from 1000 C.E. onward (Fig. S7). Finally, we note that, by construction, the LIM mean state is zero, and all anomalies decay with time. For example, when initializing a deterministic (no-noise) LIM forecast from the CCSM4-LIM reconstructed states, all global average TAS anomalies decay to nearly zero on the order of a decade (not shown). Taken together, these results strongly suggest that the relatively cool reconstructed states are a result of system memory that is consistently reinforced by information from proxies.

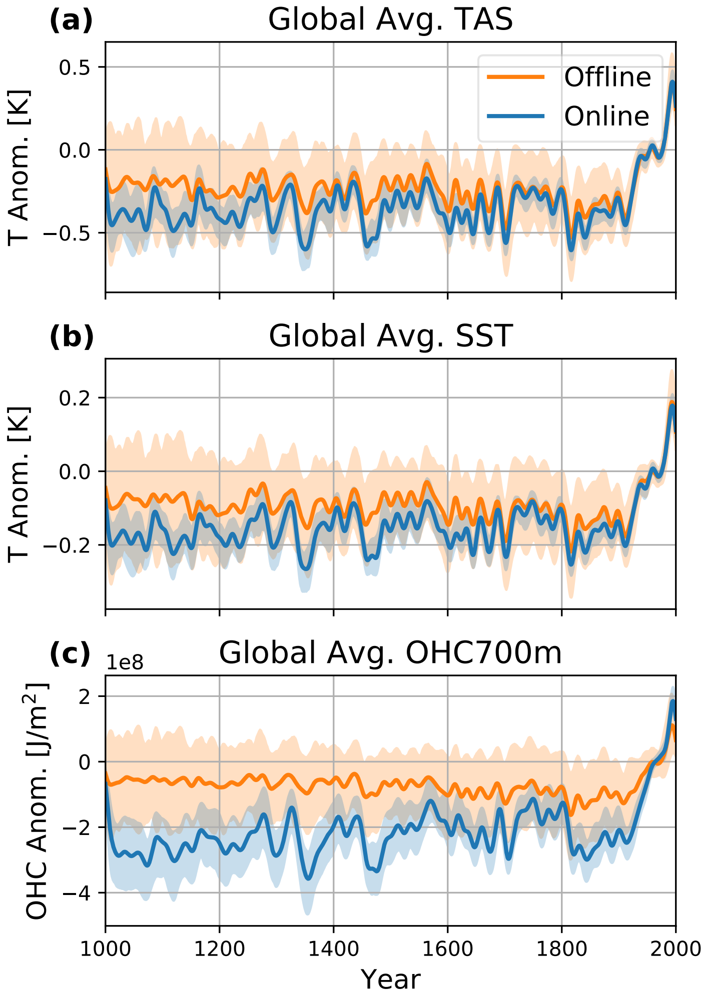

Figure 10. A comparison of the offline and online CCSM4-LIM reconstructions for lowpass filtered global average 2 m surface air temperature (TAS), sea-surface temperature (SST), and 0-700m ocean heat content (OHC700m).

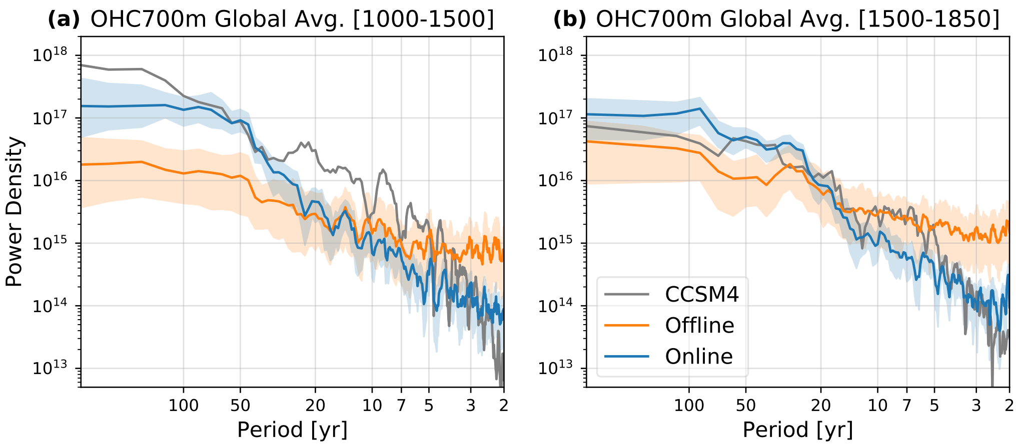

Figure 11. Spectral power density comparison of global average OHC700m between LMR online (CCSM4-LIM) (blue) and offline (orange) reconstructions and reference last millennium simulation data (dark grey) for (a) 1000–1500 C.E. and (b) 1500–1850 C.E. Spectra are calculated from the ensemble mean of each of the 50 reconstruction Monte-Carlo iterations. Solid lines denote the average spectral density while shading shows the 95% confidence interval.

The most notable effect from the inclusion of LIM forecasts for online assimilation is the memory and variability of ocean heat content. Figure 11 shows a spectral power disparity between the offline and online reconstructions at short and long periods. At short timescales (periods of 2–3 years), the offline reconstruction shows approximately an order of magnitude larger variability than the online reconstruction and reference CCSM4 simulation. At longer timescales (periods $>$ 50 years), the offline case has an order of magnitude less variability than the online reconstruction and CCSM4 reference simulation for OHC700m. This behavior is displayed in both the relatively data-sparse early period (1000–1500 C.E.) and when more proxy information is available. In the offline case, the upper-ocean heat content is solely determined by field covariance and the year's available proxy observations. The lack of memory means that OHC700m is free to vary widely between years, but also that it does not necessarily act as a long-term filter of atmospheric variability (i.e., an ocean layer with a large amount of thermal inertia). The ensemble average of global mean OHC700m (Fig. 10c) shows the lack of field constraint in the offline reconstruction. The wide confidence interval relates to a large range of reconstructed global mean OHC700m in ensemble members, but when averaged across the ensemble, little coherent low-frequency variability remains. The autocorrelation of the global average OHC700m (Figs. S9a, S9b) highlights the lack of memory in the offline reconstruction, where autocorrelations at a two-year lag decrease below 0.4, whereas the online reconstructions and reference simulations show autocorrelations in the range of 0.8–0.9.

We turn now to how the online method affects the relationships between global average SST and OHC700m (Fig. 12a,b), and the Niño 3.4 and PDO indices (Fig. 12c,d). The global average SST and OHC lead–lag correlations show online reconstructions have an asymmetric relationship where SST leads OHC700m. During the data-sparse period (Fig. 12a), the global average SST correlation with OHC700m at five-years lead is around 0.7, while at five-years lag, the correlations are around 0.4–0.5. The lead–lag relationship from the reference Last Millennium simulations is also asymmetric but with smaller correlations during this period. SST leading OHC700m implies the atmosphere is driving the changes in upper-ocean heat content. The offline reconstruction does not display

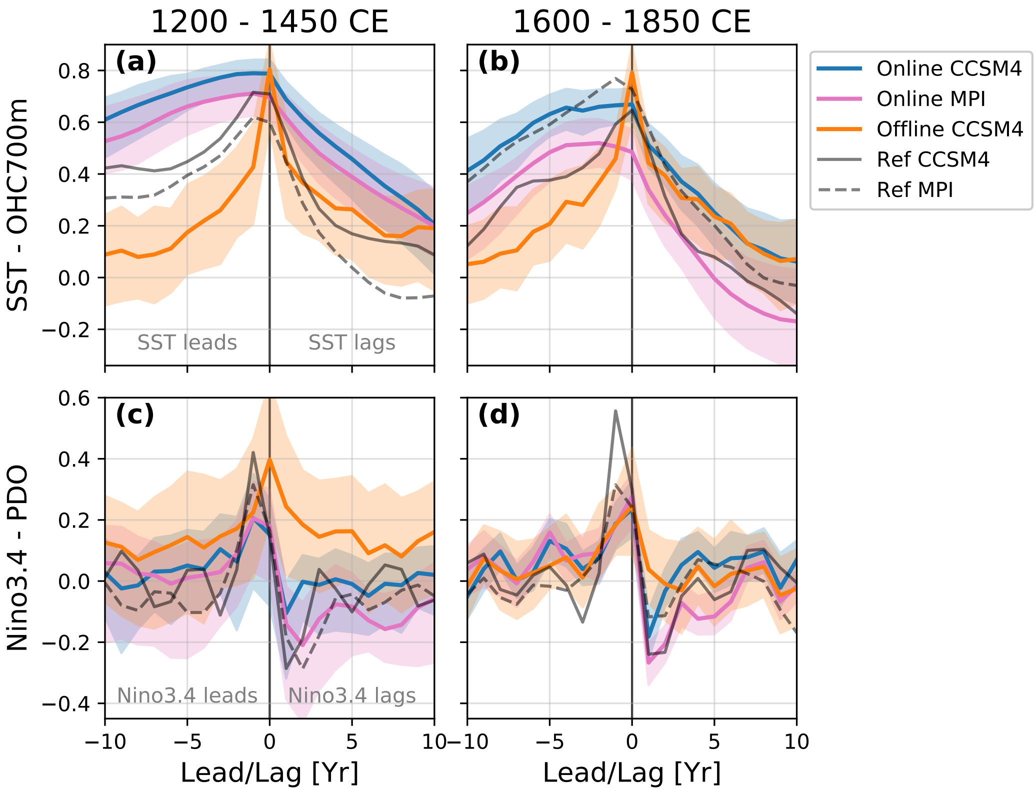

Figure 12. Lead–lag correlations for scalar indices of global average SST and OHC700m (row a) and Niño 3.4 and PDO (row c) for the time periods of 1200–1450 C.E. (column a) and 1600–1850 C.E (column b). Correlations are shown for the Online CCSM4-LIM (blue) and MPI-LIM (pink) reconstructions, the CCSM4 Offline (orange) reconstruction, and reference last millennium simulation data from the CCSM4 (black solid) and MPI (dashed black) models. Correlations are calculated at the specified lead/lag from the ensemble mean of each of the 50 reconstruction Monte-Carlo iterations. For reconstructions, solid lines denote the average correlation across Monte-Carlo iterations while shading shows the 95% confidence interval.

a lead–lag asymmetry and has lower correlation compared to the online reconstructions except at zero-lag.

From 1600–1850 C.E., when more proxy information is available, the online reconstruction lead–lag correlations for global mean SST and OHC700m are slightly lower (Fig. 12b), with correlations around 0.5–0.6 at an SST-lead of five years, and correlations near 0 (MPI-LIM) and 0.3 (CCSM4-LIM) at an SST-lag of 5 years. The online reconstructions still display a correlation asymmetry with higher correlations when SST leads. However, the online SST-lag correlations for the CCSM4-LIM reconstruction closely correspond to the SST-lag correlations in the offline case, and the MPI-LIM SST-lag correlations are lower than the offline SST-lag correlations. In the offline case, the lead–lag relationship does not qualitatively change between the two periods. The reference simulation lead–lag correlations show differences in magnitude between the periods but retain the asymmetric lead–lag character.

For the Niño 3.4 and PDO index lead–lag relationships, there is a disparity in the early period between offline and online reconstructions (Fig. 12c) that is rectified, to a degree, with increased proxy record availability (Fig. 12d). During 1200–1450 C.E., the online reconstructions show the Niño 3.4 leads the PDO by one year (correlations near 0.2). At a one-year lag, the Niño 3.4 index switches to small anticorrelation. The flip in correlations relates to the timescale difference between ENSO, which oscillates on interannual timescales, and the PDO, a more persistent decadal-scale phenomenon. The online reconstruction behavior corresponds to a similar lead–lag relationship evident in the reference simulations. The offline reconstruction again shows a symmetric relationship between the indices about zero lag, which is quite different from the online reconstructions.

During 1600–1850 C.E., when proxy availability increases, both the offline and online reconstructions are quite similar in their Niño 3.4 and PDO lead–lag correlation relationships (Fig. 12d), with the largest correlation at zero-lag. A comparison of lead–lag relationships during the instrumental period (∼1900–2000 C.E.) also show Niño 3.4 and the PDO with the highest correlation occurring at zero-lag, including for the observation-based products (Fig. S10b). The Niño 3.4 phase relationship with the PDO pattern in the North Pacific is suggested to occur via teleconnection responses in the atmospheric circulation (e.g., see review by Newman et al. (2016)). However, recent work isolating variability at different time-frequencies suggests that ENSO and the PDO may be largely independent (Wills et al., 2018). These results suggest the one-year Niño 3.4 lead is a byproduct of the GCM-calibrated LIM and less available proxy data during 1200–1450 C.E. When more proxy information is assimilated, the reconstructed PDO–ENSO relationship is in better agreement with observational products. The offline reconstruction does not show the same 1-year-lag anticorrelation as the online results, which suggests that the offline reconstruction constrained only by proxy data may be less likely to switch into La Niña conditions following the positive phase. Overall, these results highlight that when enough data are available to constrain the reconstruction, it is possible that the offline case displays the same temporal dynamics as the online method.

7 Discussion of novel reconstruction results

Our reconstructions provides novel insight into atmosphere–ocean climate fields over the last millennium based on online data assimilation of paleoclimate proxies. Here we discuss aspects in these results that differ from findings in previous studies. A notable difference compared to previous reconstructions is the finding that the LMR online reconstructions are generally colder by ∼0.25–0.5 K for NH average TAS and global average SST during the early portion of the last millennium. Additionally, our reconstructions do not show signs of a hemispheric or global-scale warm anomaly during the MCA period, and only a slight millennium-scale cooling trend into the 1800s. The lack of long-term cooling in the LMR online reconstructions contrasts with other temperature reconstructions and climate model simulations over the last millennium (e.g., Figs 8c-f). The small cooling trend with previous statistical reconstruction results should be considered within the context of known sensitivity to methodological choices (e.g., Juckes et al., 2007; Mann et al., 2008), and generally broad uncertainty bounds of previous reconstructions (see Fig. 1a in Neukom, Barboza, et al., 2019). Specifically considering the MCA–LIA differences, the reconstructed pattern of cool tropics and warmer high-latitude regions is in constrast with the primarily global pattern evident in the GCMs. We note that when removing a constant from the the reference GCM results, consistent with the colder global average of the reconstructions, the MCA–LIA difference is more similar. Our sensitivity experiments and the damped-eigenmodes of the LIM forecast (by construction) all point toward the relatively cold reconstructed early temperatures being a result of the dynamical propagation of assimilated proxy information. However, a question for future research concerns whether the reconstructed cool early period is related to the proxy distribution or lack of proxies sensitive to decadal–centennial climate variability.

Despite the relatively cold conditions of reconstructed NH average temperature from 1000–1250 C.E., the average reconstructed EU temperature displays decadal-scale warm periods, which are approximately half the magnitude of early 20th Century warming. Reconstructed spatial patterns of temperature change between the MCA and LIA period are spatially heterogeneous, showing large positive anomalies over the Barents Sea and Northern Europe and cooler tropical temperatures. The spatial character of temperature differences supports previous results depicting high-latitude warmth and cold tropical temperatures (e.g., Cobb et al., 2003; Mann et al., 2009; Goosse, Crespin, et al., 2012). Examining field regression relationships with average EU temperatures, we find that warm conditions relate to warm SST and OHC700m anomalies over the Norwegian and Barents Seas, and the mid-latitude North Atlantic. Additionally, the atmospheric circulation is characterized by lower SLP across the Arctic and increased mid-latitude ZG500 heights with an upper-level ridge centered over Europe. The reconstruction regression relationships notably differ from those derived from the reference simulations (Figs 9c–f) with more broad-scale pan-Arctic circulation connections found in the reconstruction. This again suggests that the assimilated proxies are adding new information about past dynamical relationships compared to unconstrained model simulations. The relation of EU warmth with enhanced lower-level circulations in the reconstruction at least partially supports previous work suggesting longer-term NAO-like circulation anomalies during the MCA (e.g., Trouet et al., 2009). However, the LMR online reconstructions also suggest major European warm events seem to occur on decadal scales instead of centennial scale during the MCA.

Instrumental validation of the reconstructed SST and OHC700m fields show an area of correlations that are consistently low- or anti-correlated in the Labrador Sea and regions just south of Greenland and Iceland. This region of weak correlation is also apparent in offline reconstructions (Hakim et al., 2016; Tardif et al., 2019). The lack of improvement with online and offline DA suggests that the climate models used as a basis for field constraints may not correctly represent the variability of this region during the instrumental period, and/or that we do not have enough proxy information to describe this region. In the case of unrepresentative model dynamics, it could mean that instrumental period covariances related to the North Atlantic are qualitatively different from the pre-industrial period, perhaps related to changes in external forcing. Further investigation into the causes of the reconstruction discrepancy in this region could provide insight into the relative roles of internal ocean dynamics and external forcing in the North Atlantic, which is an active and open research question (e.g., Clement et al., 2015; Zhang et al., 2016; Vecchi et al., 2017; Sutton et al., 2018; Wills et al., 2019).

8 Conclusions

Understanding the dynamics of coupled atmosphere–ocean low frequency variability of the real climate system requires a physically consistent gridded dataset that is faithful, within error, to the climate recorded by proxy archives. We have presented coupled atmosphere–ocean field reconstructions over the last millennium that incorporate temporal constraints from online data assimilation with linear inverse models (LIMs). Global aggregate measures of sea surface temperature (SST) and upper-700m ocean heat content (OHC700m) show relatively cold pre-industrial conditions with pronounced decadal-to-centennial-scale variability. By the modern era, significant increases in temperature and heat content related to anthropogenic greenhouse gas forcing are apparent. As full-field reconstructions, we can calculate and assess a large variety of physically-consistent dynamic field measures and uncertainty over a much longer time period than instrumental data allow. For example, the reconstructed Niño 3.4 index shows millennium-scale warming on the order of $\sim$0.5 K per 1000 years. The reconstructed PDO does not display distinct trends or changes to variability from 1600–2000 C.E. Before 1600, there is pronounced decadal-to-centennial-scale variability of the PDO index, potentially related to lower proxy coverage during the early period. We also find regional decadal periods of warm temperatures over northern Europe during the MCA, and from the underlying full-field information, we assess connections to regional circulation via regression against reconstructed atmospheric fields.

Instrumental validation of reconstructed ocean fields shows remarkable agreement given that the results derive from the assimilation of sparsely distributed and mostly terrestrial proxy information. We find high levels of agreement between our reconstruction and instrumental products with global average SST (correlations between 0.8–0.9), upper-700m ocean heat content (OHC700m; correlations greater than 0.9 with two recent products), and Niño 3.4 (correlations near 0.8). The reconstructed PDO shows moderate agreement with instrumental products in the CCSM4-LIM reconstruction (correlations near 0.6), generally capturing the same inter-decadal PDO variability. Furthermore, spatial validation shows broadly positive correlations for the reconstructed SST field and more regionally dependent positive correlations for OHC700m. Encouragingly, we find that our reconstructed surface temperature agrees well with the carefully curated and minimally pre-processed Crowley et al. (2014) reconstruction, and that our reconstruction gives a better estimate of the instrumental surface temperature despite the Crowley reconstruction being calibrated directly to it. Altogether, the positive validation results across many fields and indices give confidence that the reconstruction strategy produces high-fidelity, dynamically-consistent results.

The primary goal of incorporating a forecast model into the LMR framework is to provide further dynamical constraints and allow for dynamical memory of proxy information over time. For global temperature fields, assimilated proxy information suggests colder conditions on average compared to reconstructions constrained only by the proxy record. For reconstructed OHC700m, the memory from the forecast model results in an order of magnitude more power to variability on timescales longer than 50 years and smooths high-frequency fluctuations, compared to the offline case. The enhanced expression of multi-decadal variability, in general, improves one of the primary criticisms of offline reconstruction results, the relatively smooth character of reconstructed global averages. Moreover, the addition of the online forecasts improves aspects of the coupled-field lead–lag relationships in the reconstructions. For global average SST and OHC700m, lead–lag correlations show that SST generally leads OHC700m, which would physically relate to OHC integrating forcing from the ocean surface over time and is in better agreement with the GCMs from which the LIMs are derived. In contrast, the offline reconstruction shows a considerably different relationship with a symmetric peak about zero-lag between global average SST and OHC700m. The Niño 3.4 and PDO lead–lag relationship highlights the addition of dynamical constraints when fewer proxies are available. In the earlier period from 1200–1450 C.E., the online reconstructions show similar characteristics as in the GCM simulations with Niño 3.4 leading the PDO by one year, whereas the offline reconstruction again shows symmetric lead–lag correlations. During later periods, when more proxy records are available, the offline reconstruction changes character with a lead–lag relationship between Niño 3.4 and PDO similar to online reconstructions, highlighting the dependence on observations without the forecast model.

This extension of the LMR framework to include online data assimilation represents a significant step forward in combining information from proxies with climate model constraints. The approach promotes easy comparison of various GCM dynamics, provides an easy pathway to update reconstructions as new model simulations become available, and accommodates new information from expanded proxy databases (e.g., Anderson et al., 2019). For example, using two different GCM-calibrated (CCSM4 and MPI) we compare and contrast reconstructed state of past climate as a check of result robustness. While we focused on validation and description of specific ocean fields, the reconstructions include other fields, which can also be used for investigations of coupled variability over the last millennium, validation against observations and climate models, and comparisons with previous reconstructions. For example, we have used the results here to show that the time period of the MCA was in fact much colder than previously estimated in the hemispheric-mean temperature and that it was a regional phenomenon over Northern Europe. Moreover, the circulation pattern we find in the reconstruction departs significantly from that in last-millennium climate model simulations, which underscores the importance of proxy records in estimating the actual climate behavior. Similarly, when compared to last-millennium climate model simulations, the reconstructions show a greater influence of SST leading changes in OHC, and a weaker relationship between the Niño 3.4 and PDO indices (Fig. 11). These findings illustrate the power of this approach: by combining proxies and online DA, we discover dynamical insights into the climate system that differ from those in the climate model. Since climate models have different expressions of coupled atmosphere–ocean variability (e.g., Branstator et al., 2012), including the proxy records is essential.

In this work, we focus on DA-based reconstructions over the Common Era, which features relatively broad proxy coverage and many coincident GCM simulations. For deeper time DA application (e.g., Tierney et al., 2020), significant non-stationarity related to external forcing complicates the GCM modeling such that further research is needed to determine how to best formulate and apply a statistical forecast method like a LIM. For the Common Era, there are a further improvements we envision, incluidng the assimilation of multi-resolution records (e.g., Steiger & Hakim, 2016) in PDA, such as the Ocean2k sediment records (McGregor et al., 2015), which would likely provide better centennial and millennial-scale constraints than from annual timescale proxies alone. Furthermore, other fields with slow-timescales of variability, such as sea ice, might be similarly improved using an online assimilation method for field reconstruction.

Appendix A EnKF Vector Solver

Here we present the vector solver for the ensemble Kalman update equation, which is a variant of the ensemble transform Kalman filter described by Bishop et al. (2002). When using ensembles to estimate sample statistics (e.g., prior covariance, $\mathbf{B}$), the Kalman update equation (Eq. 1) describes an update of the ensemble average. The state, $\mathbf{x}$, is an $M \times 1$ row vector equal to the the column average of, $\mathbf{X}$, the $M \times N$ state ensemble. Here, M represents the number of state features (e.g., field grid points) and N is the number of ensemble members. Equivalently, the estimated observations, $\mathbf{y}_e$ (with dimensions $P \times 1$ where P is the number of proxies) represent the column average of the ensemble of proxy estimates calculated on $\mathcal{H}(\mathbf{X})$. For paleoclimate reconstruction, it is usually the case that the number of state features is much larger than the number of observations ($M \gg P$), which suggests that the updated state will be at least partially under-determined. The vector transform reduces the problem into the smaller space ($O(N\times P)$) by diagonalizing $\mathbf{K}$ and creating linear combinations of proxy observations.

To translate the problem into a transformed space, we define $\widetilde{\mathbf{x}} = \mathbf{B}^{-1/2}\mathbf{x}$ and $\widetilde{\mathbf{y}} = \mathbf{R}^{-1/2}\mathbf{y}$ and substitute into Eq. 1, giving

$$\widetilde{\mathbf{x}}_a = \widetilde{\mathbf{x}}_b + \widetilde{\mathbf{K}}[\widetilde{\mathbf{y}} - \widetilde{\mathbf{H}}\widetilde{\mathbf{x}}_b] \tag{A1}$$

where

$$\begin{aligned} \mathbf{B}^{1/2} &= \frac{1}{\sqrt{N-1}}[\mathbf{X} - \overline{\mathbf{X}}], \\ \widetilde{\mathbf{H}} &= \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}, \\ \widetilde{\mathbf{K}} &= \widetilde{\mathbf{H}}^T \left[\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T + \mathbf{I}\right]^{-1}. \end{aligned}$$

We then perform a singular value decomposition on $\widetilde{\mathbf{H}}$ ($\widetilde{\mathbf{H}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$) and transform into the new component space using

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{V}^T\widetilde{\mathbf{x}} \\ \hat{\mathbf{y}} &= \mathbf{U}^T\widetilde{\mathbf{y}}. \end{aligned}$$

Converting equation A1 into the new coordinates yields

$$\hat{\mathbf{x}}_a = \hat{\mathbf{x}}_b + \hat{\mathbf{K}}[\hat{\mathbf{y}} - \mathbf{\Lambda}\hat{\mathbf{x}}_b], \tag{A2}$$

where

$$\hat{\mathbf{K}} = \mathbf{\Lambda}\left[\mathbf{\Lambda}^2 + I\right]^{-1}, \tag{A3}$$

In this optimal space, the update considers all proxies and properly weights information between ensemble members in a single calculation. The weighting term, $\mathbf{\Lambda}$ ($P\times N$), is a diagonal matrix with terms in order of influence. The optimal space state ($\hat{\mathbf{x}}$) has dimensions of $N \times 1$ and observations ($\hat{\mathbf{y}}$) has dimensions $P \times 1$. To translate from the optimal space state back to the full space for results, we calculate

$$\mathbf{X}_a = \mathbf{X}_b\mathbf{V}\hat{\mathbf{x}}_a. \tag{A4}$$

Appendix B Linear Inverse Models

For the benefit of the reader, we summarize a technical background description of LIMs, their calibration, and stochastic integration in Appendix B1, and in Appendix B2, we summarize the multivariate LIM calibration strategy described in Perkins and Hakim (2020).

B1 Background

A linear inverse model (LIM, Penland & Sardeshmukh, 1995) is an empirically determined estimate of a dynamical system linearized about its mean state. In this model,

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \boldsymbol{\xi}, \tag{B1}$$

the slow-varying deterministic drift of the state is explicitly defined by the matrix operator, $\mathbf{L}$, while fast-timescale processes are represented as white-noise forcing, $\boldsymbol{\xi}$. We use this simple model to create a coupled GCM analog to forecast multivariate states between reconstructed times.

To create a LIM, we perform an empirical fit based on 1-year lag-covariance ($\tau = 1$) statistics of the climate state,

$$\mathbf{L} = \tau^{-1}\ln[\mathbf{C}(\tau)\mathbf{C}(0)^{-1}], \tag{B2}$$

with the sample n-lag covariance defined as $\mathbf{C}(n) = \langle \mathbf{x}(n)\mathbf{x}^T(0) \rangle$. Note that angle brackets represent an expectation, which in practice is taken as a sample average. The dynamical operator, $\mathbf{L}$, encapsulates the information to propagate predictable aspects of the state from one time to the next. System dynamics are assumed to be stable, which requires that the forecast modes (i.e., eigenvectors of $\mathbf{L}$) of a valid LIM do not grow with time. While constructive interference between forecast modes enables short-term transient anomaly growth (e.g., ENSO), extended deterministic forecasts asymptote to zero if all modes are damped. For applications in EnKF assimilation, the forecast ensemble variance is crucial for weighting the prior against and the innovation from observations. LIM deterministic forecasts alone (omitting the noise term, $\boldsymbol{\xi}$) have collapsing forecast ensemble variance over time, which limits the utilization of observations.

In Perkins and Hakim (2017), forecast ensemble variance is enhanced by blending the LIM ensemble forecast with climatological covariances from a GCM, using a "hybrid" DA method adapted from Hamill and Snyder (2000). This technique incorporates a blending coefficient, which is tuned based on reconstruction results, to control the amount of information from the forecast and climatological source. More recently, PH20 use a LIM as a GCM analog for coupled ocean–atmosphere ensemble forecasts and find that stochastic LIM forecasts reasonably approximate the ensemble variance and errors at 1-year lead times. Stochastic integration provides a natural mechanism to sample the envelope of noise-forced spread across assimilation times, and it does not necessarily require a tuning procedure based on reconstruction output. Moreover, integration provides a straightforward technique for providing time-variable information that is potentially useful for more sophisticated PSMs in the future. For these reasons, we choose to utilize a LIM with stochastic noise forcing ($\boldsymbol{\xi}$) to sustain ensemble variance between assimilation times.

To perform stochastic LIM forecasts, we first determine the noise forcing statistics of the system, $\mathbf{Q} = \langle \boldsymbol{\xi}\boldsymbol{\xi}^T \rangle dt$, using the calibration data. With the assumption of stationary statistics, we use the dynamical operator, $\mathbf{L}$ (Eq. B2), and the fluctuation–dissipation relationship (Penland & Matrosova, 1994),

$$\frac{d\mathbf{C}(0)}{dt} = \mathbf{L}\mathbf{C}(0) + \mathbf{C}(0)\mathbf{L}^T + \mathbf{Q} = 0, \tag{B3}$$

to estimate $\mathbf{Q}$. With both model terms ($\mathbf{L}$ and $\mathbf{Q}$) we explicitly simulate the deterministic and stochastic drift of a sample climate trajectory over time using a two-step integration scheme defined by Penland and Matrosova (1994),

$$\mathbf{a}(t+\delta t) = \mathbf{L}\mathbf{x}(t) + \hat{\mathbf{Q}}\sqrt{\boldsymbol{\Lambda}\delta t}\,\boldsymbol{\alpha} \tag{B4}$$

$$\mathbf{x}(t+\delta t/2) = [\mathbf{x}(t) + \mathbf{a}(t+\delta t)]/2. \tag{B5}$$

In this scheme, $\mathbf{a}$ is an intermediate state variable, $\hat{\mathbf{Q}}$ and $\mathbf{\Lambda}$ are from the eigendecomposition $\mathbf{Q} = \hat{\mathbf{Q}}\mathbf{\Lambda}\hat{\mathbf{Q}}^{-1}$, δt is the timestep, and $\boldsymbol{\alpha}$ is a vector of random numbers drawn from a unit-normal distribution.

B2 Parameter Reduction and Calibration

Climate-scale predictability is typically dominated by a few modes of atmosphere–ocean variability such as ENSO, the PDO, and Atlantic multidecadal variability. Therefore, the number of important degrees of freedom for the predictable climate state are substantially fewer than the total degrees of freedom presented by the gridded climate fields of the climate model. For this reason, we consolidate the information of the gridded climate fields for use in LIM forecasting by using a two-step EOF reduction as in PH20, which we summarize here.

The climate state, $\mathbf{X}$, used for reconstruction includes target output fields and also fields required for estimated observation calculations via PSMs. These fields are concatenated along the first dimension to form the total state,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_F \end{bmatrix}.$$

For the initial step, we reduce each of the F fields individually. We first area weight each row of field $\mathbf{X}_f$ (with dimensions of spatial features by ensemble samples, $m \times n$), in this case using latitude weighting,

$$\mathbf{x}_i^w = \mathbf{x}_i \sqrt{cos(\phi_i)} \quad \text{for}\, i = 1,\, ...,\, m,$$

where ϕ_i represents the i^{th} grid cell latitude. We then find the EOFs truncated to the leading k modes, $\mathbf{U}_f$, using a singular value decomposition (SVD), $\mathbf{X}_f^w = \mathbf{U}_f \mathbf{\Sigma}_f \mathbf{V}_f^T$, and project the field into this space using: $\hat{\mathbf{X}}_f = \mathbf{U}_f^T \mathbf{X}_f$. The initial reduction compactly represents fields and expedites the next step's multivariate-EOF calculation.

For the second reduction, we first standardize each field by the total component variance, σ_f, and reform the state using previously reduced fields,

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{\mathbf{X}}_1/\sigma_1 \\ \hat{\mathbf{X}}_2/\sigma_2 \\ \vdots \\ \hat{\mathbf{X}}_F/\sigma_F \end{bmatrix}.$$

Then we take the SVD of this new state, $\hat{\mathbf{X}} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^T$, and use the leading ℓ EOFs to project into the components of the fully reduced multivariate-EOF state, $\widetilde{\mathbf{X}} = \hat{\mathbf{U}}^T\hat{\mathbf{X}}$. The LIM calibration procedure (Eq. B2) and forecasts (Eqs. B4,B5) take place using data projected into the reduced space defined by $\widetilde{\mathbf{X}}$. By storing the EOFs and standardization factors ($\hat{\mathbf{U}}$, $\mathbf{U}_f$, and σ_f), we project into and out of the space during the reconstruction process.

The reduction process introduces two parameters, k and ℓ, determining the EOF truncation at each step. For our experiments, we choose to retain the leading 400 EOFs ($k = 400$) in the first step, which we find retains greater than 90% of each field's variance. The second reduction parameter ultimately determines the LIM skill properties and ensemble forecast spread characteristics (Perkins & Hakim, 2020). As in PH20, we separate OHC700m from the multivariate reduction process due to adverse effects on forecast skill from truncation. The leading OHC700m field components from the first reduction (retaining 20 modes) are instead appended to the multivariate state to form the LIM-space components,

$$\mathbf{Z} = \begin{bmatrix} \widetilde{\mathbf{X}} \\ \hat{\mathbf{X}}_{OHC} \end{bmatrix}.$$

We additionally find that due to the small magnitude of values in the PR field, which have similar scale as numerical errors of the EOF-reduction procedure, it is useful to standardize precipitation values by the total field variance before the first step reduction.

With additional fields in the state vector compared to PH20, including all seasonally averaged TAS fields, we run the same diagnostic tests for increasing numbers of retained multivariate-EOF components investigating both aggregate measures of 1-year forecast skill and ensemble calibration. We get similar results as in PH20 that the forecast skill of aggregate measures (e.g., global averages, ENSO, and the PDO) are positive and do not substantially change after retaining 15–20 multivariate-EOF components (not shown). Ensemble characteristics are prone to unpredictable changes at different multivariate-EOF truncations due to an imperfectly determined $\mathbf{L}$ (e.g., from noise and non-linearity), which similarly affects the determination of $\mathbf{Q}$ and leads to numerical uncertainty in the eigendecomposition of the noise statistics (PH20). Therefore, we perform tests to assess the ensemble calibration ratios (see supplementary text Section S1) for global averages (TAS, SST, OHC700m) and dynamic indices (PDO- and ENSO-related quantities). The calibration ratio gives a measure indicating whether the ensemble variance is representative of the forecast errors where a “well-calibrated” ensemble forecast system would result in a ratio near 1.0. Based on calibration ratio results, we select the multivariate-EOF truncation parameter of $\ell = 20$ for the CCSM4-LIM, and $\ell = 27$ for the MPI-LIM.

Appendix C PSM Objective Seasonality Determination

For proxy PSMs, we use univariate linear regression models, $y_{ek} = \beta_{0k} + \beta_{1k}\overline{x} + \epsilon_k$, fit to temperature data estimate proxy values for the k^{th} proxy, y_{ek} (e.g., tree-ring widths), from the climate state, $\overline{x}$. The overbar (e.g., $\overline{x}$) denotes a seasonal-to-annual average of the temperature using data from the closest grid cell from calibration data. We determine the time-average distinction used for each PSM by a series of objective seasonality tests for tree-based proxies (as in Tardif et al., 2019), and use the PAGES2017 seasonality metadata for all other proxies. We fit parameters including the intercept (β_{0k}), slope (β_{1k}) and Gaussian error with statistics $\mathcal{N}(0, \sigma_k^2)$ through a least-squares fit with co-located temperature data in the NASA Goddard Institute for Space Studies Surface Temperature Analysis GISTEMP v4 (Hansen et al., 2010) dataset (similarly regridded to a 2° x 2° grid). Note that the proxy error variances (σ_k^2), which are assumed to be independent, form the diagonal of the proxy error covariance matrix, $\mathbf{R}$. The objective determination of seasonality for tree-based proxies involves a series of tests across prescribed seasonal averages (Jan–Dec, JJA, JJASON, DJF, DJFMAM, AMJJAS, OND-JFM) and the expert-derived seasonality in the PAGES2017 database. We then make a selection of the seasonal-average definition with the best overall calibration fit from the tests. The best fit is defined as the model with the lowest-value Bayesian information criterion, BIC $= -2\ln(\hat{L}) + k\ln(n)$ (Schwarz, 1978), where $\hat{L}$ is the maximum value from the likelihood function of the model, k is the model's number of estimated parameters, and n is the sample size.

Acronyms

CCSM4 Community Climate System Model version 4
CCSM4-LIM CCSM4-calibrated LIM
CFR climate field reconstruction
DA data assimilation
EG13 Emile-Geay et al. (2013a)
EnKF ensemble Kalman filter
EOF empirical orthogonal function
GCM global climate model
LIA Little Ice Age
LIM linear inverse model
LMR Last Millennium Reanalysis
MCA Medieval Climate Anomaly
MPI Max Planck Institute Earth System Model
MPI-LIM MPI-calibrated LIM
OHC700m 0–700 m ocean heat content
PAGES2017 PAGES 2k Consortium (2017)
PDA paleoclimate data assimilation
PDO Pacific Decadal Oscillation
PH20 Perkins and Hakim (2020)
PR precipitation
PSM proxy system model
RLUT outgoing TOA longwave radiation
RSUT outgoing TOA shortwave radiation
SLP sea-level pressure
SSS sea-surface salinity
SST sea-surface temperature
TAS 2 m surface air temperature
TOA top of atmosphere
ZG500 500 hPa geopotential heights
ZOS dynamic ocean surface height

Acknowledgments

The authors thank Robert Tardif for his contributions to the LMR codebase, many of which are adapted for this online version. Chris Snyder (NCAR) originated the idea for the vector solver described in Appendix A. The comparison of our results to Crowley et al. (2014) was suggested by an anonymous reviewer and the Associate Editor. We thank three anonymous reviewers, the Associate Editor, and the Editor for their feedback on the manuscript. This research was supported by Heising-Simons Foundation grant 2016-014, and National Science Foundation (NSF) grants AGS-1602223, AGS-1702423 made by the to the University of Washington. The first author was supported by the National Science Foundation Graduate Research Fellowship under grant no. DGE-1256082.

The CMIP5 climate model data was obtained through the Lawrence Livermore National Laboratory Earth System Grid Federation Portal (https://esgf-node.llnl.gov/projects/esgf-llnl/). The reconstruction framework used to perform the experiments can be found online at https://github.com/frodre/LMROnline (https://doi.org/10.5281/zenodo.4563364). The linear inverse modeling code can be found at

https://github.com/frodre/pyLIM (http://doi.org/10.5281/zenodo.3243749). Selected reconstruction data, analysis code, and pre-processed LMR input data are available at http://doi.org/10.5281/zenodo.4563663.

References

Alexander, M. A., Matrosova, L., Penland, C., Scott, J. D., & Chang, P. (2008, jan). Forecasting Pacific SSTs: Linear Inverse Model Predictions of the PDO. *Journal of Climate*, *21*(2), 385–402. doi: 10.1175/2007JCLI1849.1

Ambaum, M. H. P., Hoskins, B. J., & Stephenson, D. B. (2001, aug). Arctic Oscillation or North Atlantic Oscillation? *Journal of Climate*, *14*(16), 3495–3507. doi: 10.1175/1520-0442(2001)014⟨3495:AOONAO⟩2.0.CO;2

Ammann, C. M., Joos, F., Schimel, D. S., Otto-Bliesner, B. L., & Tomas, R. A. (2007). Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model. *Proceedings of the National Academy of Sciences*, *104*(10), 3713–3718. doi: 10.1073/pnas.0605064103

Anderson, D. M., Tardif, R., Horlick, K., Erb, M. P., Hakim, G. J., Noone, D., ... Steig, E. (2019). Additions to the Last Millennium Reanalysis Multi-Proxy Database. *Data Science Journal*, *18*(1), 2. doi: 10.5334/dsj-2019-002

Atwood, A. R., Wu, E., Frierson, D. M. W., Battisti, D. S., & Sachs, J. P. (2016, feb). Quantifying Climate Forcings and Feedbacks over the Last Millennium in the CMIP5–PMIP3 Models*. *Journal of Climate*, *29*(3), 1161–1178. doi: 10.1175/JCLI-D-15-0063.1

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1132–1161. doi: 10.1002/qj.2063

Bhend, J., Franke, J., Folini, D., Wild, M., & Brönnimann, S. (2012, may). An ensemble-based approach to climate reconstructions. *Climate of the Past*, *8*(3), 963–976. doi: 10.5194/cp-8-963-2012

Biondi, F., Gershunov, A., & Cayan, D. R. (2001). North Pacific Decadal Climate Variability since 1661. *Journal of Climate*, *14*(1), 5–10. doi: 10.1175/1520-0442(2001)014⟨0005:NPDCVS⟩2.0.CO;2

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2002). Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, *129*(3), 420–436. doi: 10.1175/1520-0493(2001)129⟨0420:aswtet⟩2.0.co;2

Branstator, G., Teng, H., Meehl, G. A., Kimoto, M., Knight, J. R., Latif, M., & Rosati, A. (2012). Systematic estimates of initial-value decadal predictability for six AOGCMs. *Journal of Climate*, *25*(6), 1827–1846. doi: 10.1175/JCLI-D-11-00227.1

Broecker, W. S. (2001, feb). PALEOCLIMATE: Was the Medieval Warm Period Global? *Science*, *291*(5508), 1497–1499. doi: 10.1126/science.291.5508.1497

Chang, Y. S., Zhang, S., Rosati, A., Delworth, T. L., & Stern, W. F. (2013). An assessment of oceanic variability for 1960-2010 from the GFDL ensemble coupled data assimilation. *Climate Dynamics*, *40*(3-4), 775–803. doi: 10.1007/s00382-012-1412-2

Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., & Zhu, J. (2017). Improved estimates of ocean heat content from 1960 to 2015. *Science Advances*, *3*(3), 1–11. doi: 10.1126/sciadv.1601545

Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Radel, G., & Stevens, B. (2015, oct). The Atlantic Multidecadal Oscillation without a role for ocean circulation. *Science*, *350*(6258), 320–324. doi: 10.1126/science.aab3980

Cobb, K. M., Charles, C. D., Cheng, H., & Edwards, R. L. (2003, jul). El

Niño/Southern Oscillation and tropical Pacific climate during the last millennium. *Nature*, *424*(6946), 271–276. doi: 10.1038/nature01779
Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., ... Worley, S. J. (2011, jan). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 1–28. doi: 10.1002/qj.776
Crespin, E., Goosse, H., Fichefet, T., & Mann, M. E. (2009). The 15th century Arctic warming in coupled model simulations with data assimilation. *Climate of the Past*, *5*(3), 389–401. doi: 10.5194/cp-5-389-2009
Crowley, T. J., Obrochta, S. P., & Liu, J. (2014, may). Recent global temperature "plateau" in the context of a new proxy reconstruction. *Earth's Future*, *2*(5), 281–294. doi: 10.1002/2013EF000216
D'Arrigo, R., Villalba, R., & Wiles, G. (2001). Tree-ring estimates of Pacific decadal climate variability. *Climate Dynamics*, *18*(3-4), 219–224. doi: 10.1007/s003820100177
D'Arrigo, R., & Wilson, R. (2006). On the Asian expression of the PDO. *International Journal of Climatology*, *26*(12), 1607–1617. doi: 10.1002/joc.1326
de Boisséson, E., Balmaseda, M. A., & Mayer, M. (2018). Ocean heat content variability in an ensemble of twentieth century ocean reanalyses. *Climate Dynamics*, *50*(9-10), 3783–3798. doi: 10.1007/s00382-017-3845-0
Diaz, H. F., Trigo, R., Hughes, M. K., Mann, M. E., Xoplaki, E., & Barriopedro, D. (2011, nov). Spatial and Temporal Characteristics of Climate in Medieval Times Revisited. *Bulletin of the American Meteorological Society*, *92*(11), 1487–1500. doi: 10.1175/BAMS-D-10-05003.1
Dubinkina, S., Goosse, H., Sallaz-Damaz, Y., Crespin, E., & Crucifix, M. (2011, dec). Testing a Particle Filter to Reconstruct Climate Changes Over the Past Centuries. *International Journal of Bifurcation and Chaos*, *21*(12), 3611–3618. doi: 10.1142/S0218127411030763
Duchon, C. E. (1979, aug). Lanczos Filtering in One and Two Dimensions. *Journal of Applied Meteorology*, *18*(8), 1016–1022. doi: 10.1175/1520-0450(1979)018⟨1016:LFIOAT⟩2.0.CO;2
Emile-Geay, J., Cobb, K. M., Mann, M. E., & Wittenberg, A. T. (2013a, apr). Estimating Central Equatorial Pacific SST Variability over the Past Millennium. Part II: Reconstructions and Implications. *Journal of Climate*, *26*(7), 2329–2352. doi: 10.1175/JCLI-D-11-00511.1
Emile-Geay, J., Cobb, K. M., Mann, M. E., & Wittenberg, A. T. (2013b, apr). Estimating Central Equatorial Pacific SST Variability over the Past Millennium. Part I: Methodology and Validation. *Journal of Climate*, *26*(7), 2302–2328. doi: 10.1175/JCLI-D-11-00510.1
Esper, J., Cook, E. R., & Schweingruber, F. H. (2002). Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science*, *295*(5563), 2250–2253. doi: 10.1126/science.1066208
Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. doi: 10.1007/s10236-003-0036-9
Franke, J., Brönnimann, S., Bhend, J., & Brugnara, Y. (2017). A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations. *Scientific Data*, *4*, 1–19. doi: 10.1038/sdata.2017.76
Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., ... Stevens, B. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597. doi: 10.1002/jame.20038
Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with

uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*(12), 6704–6716. doi: 10.1002/2013JC009067

Goosse, H. (2017). Reconstructed and simulated temperature asymmetry between continents in both hemispheres over the last centuries. *Climate Dynamics*, *48*(5-6), 1483–1501. doi: 10.1007/s00382-016-3154-z

Goosse, H., Crespin, E., de Montety, a., Mann, M. E., Renssen, H., & Timmermann, a. (2010). Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *Journal of Geophysical Research*, *115*, 1–17. doi: 10.1029/2009JD012737

Goosse, H., Crespin, E., Dubinkina, S., Loutre, M. F., Mann, M. E., Renssen, H., ... Shindell, D. (2012). The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly". *Climate Dynamics*, *39*(12), 2847–2866. doi: 10.1007/s00382-012-1297-0

Goosse, H., Guiot, J., Mann, M. E., Dubinkina, S., & Sallaz-Damaz, Y. (2012). The medieval climate anomaly in Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation. *Global and Planetary Change*, *84-85*, 35–47. doi: 10.1016/j.gloplacha.2011.07.002

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., ... Perkins, W. A. (2016). The Last Millennium Climate Reanalysis Project: Framework and First Results. *Journal of Geophysical Research: Atmospheres*. doi: 10.1002/2016JD024751

Hamill, T. M., & Snyder, C. (2000). A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Monthly Weather Review*, *128*(8), 2905–2919. doi: 10.1175/1520-0493(2000)128⟨2905:AHEKFV⟩2.0.CO;2

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Rev. Geophys.*, *48*(4), RG4004. doi: 10.1029/2010RG000345.1.INTRODUCTION

Hasselmann, K. (1976). Stochastic climate models Part I. Theory. *Tellus*, *28*(6), 473–485. doi: 10.3402/tellusa.v28i6.11316

Hawkins, E., & Sutton, R. (2009). Decadal predictability of the Atlantic Ocean in a coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling. *Journal of Climate*, *22*(14), 3960–3978. doi: 10.1175/2009JCLI2720.1

Huntley, H. S., & Hakim, G. J. (2010, nov). Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Climate Dynamics*, *35*(6), 995–1009. doi: 10.1007/s00382-009-0714-5

Juckes, M. N., Allen, M. R., Briffa, K. R., Esper, J., Hegerl, G. C., Moberg, A., ... Weber, S. L. (2007, oct). Millennial temperature reconstruction intercomparison and evaluation. *Climate of the Past*, *3*(4), 591–609. doi: 10.5194/cp-3-591-2007

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J. R., Broennimann, S., Buizza, R., ... Schepers, D. (2018). CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, *10*(5), 1172–1195. doi: 10.1029/2018MS001273

Lamb, H. H. (1965). The early medieval warm epoch and its sequel. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *1*, 13–37.

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., & Teng, H. (2013). Last Millennium Climate and Its Variability in CCSM4. *Journal of Climate*, *26*(4), 1085–1111. doi: 10.1175/JCLI-D-11-00326.1

Ljungqvist, F. C. (2010, sep). A new reconstruction of temperature variability in the extra-tropical northern hemisphere during the last two millennia. *Geografiska Annaler: Series A, Physical Geography*, *92*(3), 339–351. doi: 10.1111/j.1468-0459.2010.00399.x

MacDonald, G. M., & Case, R. A. (2005). Variations in the Pacific Decadal Oscillation over the past millennium. *Geophysical Research Letters*, *32*(8), 1–4. doi: 10.1029/2005GL022478

Mann, M. E., Bradley, R. S., & Hughes, M. K. (1999). *Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations* (Vol. 26). doi: 10.1029/1999GL900070

Mann, M. E., & Jones, P. D. (2003). Global surface temperatures over the past two millennia. *Geophysical Research Letters*, *30*(15), 15–18. doi: 10.1029/2003GL017814

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., & Ni, F. (2008). Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences*, *105*(36), 13252–13257. doi: 10.1073/pnas.0805721105

Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., ... Ni, F. (2009, nov). Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly. *Science*, *326*(5957), 1256–1260. doi: 10.1126/science.1177303

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. *Bulletin of the American Meteorological Society*, *78*(6), 1069–1079. doi: 10.1175/1520-0477(1997)078⟨1069:APICOW⟩2.0.CO;2

Matsikaris, A., Widmann, M., & Jungclaus, J. (2015, jan). On-line and off-line data assimilation in palaeoclimatology: a case study. *Climate of the Past*, *11*(1), 81–93. doi: 10.5194/cp-11-81-2015

Matsikaris, A., Widmann, M., & Jungclaus, J. (2016a, jun). Assimilating continental mean temperatures to reconstruct the climate of the late pre-industrial period. *Climate Dynamics*, *46*(11-12), 3547–3566. doi: 10.1007/s00382-015-2785-9

Matsikaris, A., Widmann, M., & Jungclaus, J. (2016b). Influence of proxy data uncertainty on data assimilation for the past climate. *Climate of the Past*, *12*(7), 1555–1563. doi: 10.5194/cp-12-1555-2016

McGregor, H. V., Evans, M. N., Goosse, H., Leduc, G., Martrat, B., Addison, J. A., ... Ersek, V. (2015). Robust global ocean cooling trend for the pre-industrial Common Era. *Nature Geoscience*, *8*(9), 671–677. doi: 10.1038/ngeo2510

Meehl, G. A., Arblaster, J. M., Matthes, K., Sassi, F., & Van Loon, H. (2009). Amplifying the pacific climate system response to a small 11-year solar cycle forcing. *Science*, *325*(5944), 1114–1118. doi: 10.1126/science.1172872

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, M. H., & Karlén, W. (2005). Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, *433*(7026), 613–617. doi: 10.1038/nature03265

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., ... Von Gunten, L. (2019, aug). Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nature Geoscience*, *12*(8), 643–649. Retrieved from `http://www.nature.com/articles/s41561-019-0400-0` doi: 10.1038/s41561-019-0400-0

Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., & Werner, J. P. (2019, jul). No evidence for globally coherent warm and cold periods over the preindustrial Common Era. *Nature*, *571*(7766), 550–554. doi: 10.1038/s41586-019-1401-2

Newman, M. (2013, jul). An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies. *Journal of Climate*, *26*(14), 5260–5269. doi: 10.1175/JCLI-D-12-00590.1

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., ... Smith, C. A. (2016). The Pacific decadal oscillation, revisited. *Journal* *of Climate*, *29*(12), 4399–4427. doi: 10.1175/JCLI-D-15-0508.1

Newman, M., Alexander, M. A., & Scott, J. D. (2011). An empirical model of tropical ocean dynamics. *Climate Dynamics*, *37*(9-10), 1823–1841. doi: 10.1007/s00382-011-1034-0

Okazaki, A., & Yoshimura, K. (2017). Development and evaluation of a system of proxy data assimilation for paleoclimate reconstruction. *Climate of the Past*, *13*(4), 379–393. doi: 10.5194/cp-13-379-2017

Oke, P. R. (2002). Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research*, *107*(C9), 3122. doi: 10.1029/2000JC000511

Oke, P. R., Sakov, P., & Corney, S. P. (2007, jan). Impacts of localisation in the EnKF and EnOI: experiments with a small model. *Ocean Dynamics*, *57*(1), 32–45. doi: 10.1007/s10236-006-0088-8

Oke, P. R., Schiller, A., Griffin, D. A., & Brassington, G. B. (2005, oct). Ensemble data assimilation for an eddy-resolving ocean model of the Australian region. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3301–3311. doi: 10.1256/qj.05.95

Osborn, T. J., Jones, P. D., Lister, D. H., Morice, C. P., Simpson, I. R., Winn, J. P., ... Harris, I. C. (2021, jan). Land Surface Air Temperature Variations Across the Globe Updated to 2019: The CRUTEM5 Data Set. *Journal of Geophysical Research: Atmospheres*, *126*(2). doi: 10.1029/2019JD032352

PAGES 2k Consortium. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*(170088), 1–33. doi: DOI: 10.1038/sdata.2017.88

Pendergrass, A. G., Hakim, G. J., Battisti, D. S., & Roe, G. (2012, jan). Coupled Air–Mixed Layer Temperature Predictability for Climate Reconstruction. *Journal of Climate*, *25*(2), 459–472. doi: 10.1175/2011JCLI4094.1

Penland, C., & Matrosova, L. (1994). *A Balance Condition for Stochastic Numerical Models with Application to the El Niño-Southern Oscillation* (Vol. 7) (No. 9). doi: 10.1175/1520-0442(1994)007⟨1352:ABCFSN⟩2.0.CO;2

Penland, C., & Sardeshmukh, P. D. (1995, aug). The Optimal Growth of Tropical Sea Surface Temperature Anomalies. *Journal of Climate*, *8*(8), 1999–2024. doi: 10.1175/1520-0442(1995)008⟨1999:TOGOTS⟩2.0.CO;2

Perkins, W. A., & Hakim, G. (2020). Linear inverse modeling for coupled atmosphere-ocean ensemble climate prediction. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001778. (e2019MS001778 10.1029/2019MS001778) doi: 10.1029/2019MS001778

Perkins, W. A., & Hakim, G. J. (2017). Reconstructing paleoclimate fields using online data assimilation with a linear inverse model. *Climate of the Past*, *13*(5), 421–436. doi: 10.5194/cp-13-421-2017

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., ... Fisher, M. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, *29*(11), 4083–4097. doi: 10.1175/JCLI-D-15-0556.1

Pollack, H. N., & Smerdon, J. E. (2004). Borehole climate reconstructions: Spatial structure and hemispheric averages. *Journal of Geophysical Research D: Atmospheres*, *109*(11), 1–9. doi: 10.1029/2003JD004163

Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbéoch, M., ... Jayne, S. R. (2016). Fifteen years of ocean observations with the global Argo array. *Nature Climate Change*, *6*(2), 145–153. doi: 10.1038/nclimate2872

Rogers, J. C. (1997). North Atlantic storm track variability and its association to the North Atlantic oscillation and climate variability of Northern Europe. *Journal of Climate*, *10*(7), 1635–1647. doi: 10.1175/1520-0442(1997)010⟨1635:NASTVA⟩2.0.CO;2

Rutherford, S., Mann, M. E., Osborn, T. J., Briffa, K. R., Jones, P. D., Bradley, R. S., & Hughes, M. K. (2005, jul). Proxy-Based Northern Hemisphere Surface

Temperature Reconstructions: Sensitivity to Method, Predictor Network, Target Season, and Target Domain. *Journal of Climate*, *18*(13), 2308–2329. doi: 10.1175/JCLI3351.1

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Shen, C., Wang, W. C., Gong, W., & Hao, Z. (2006). A Pacific Decadal Oscillation record since 1470 AD reconstructed from proxy data of summer rainfall over eastern China. *Geophysical Research Letters*, *33*(3), 1–4. doi: 10.1029/2005GL024804

Singh, H. K. A., Hakim, G. J., Tardif, R., Emile-Geay, J., & Noone, D. C. (2017). Atlantic Multidecadal Variability from the Last Millennium Reanalysis. *Climate of the Past Discussions*, 1–25. doi: 10.5194/cp-2017-49

Steiger, N. J., & Hakim, G. (2016, jun). Multi-timescale data assimilation for atmosphere–ocean state estimates. *Climate of the Past*, *12*(6), 1375–1388. doi: 10.5194/cp-12-1375-2016

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., & Roe, G. H. (2014). Assimilation of Time-Averaged Pseudoproxies for Climate Reconstruction. *Journal of Climate*, *27*, 426–441. doi: 10.1175/JCLI-D-12-00693.1

Steiger, N. J., Smerdon, J. E., Cook, E. R., & Cook, B. I. (2018). A reconstruction of global hydroclimate and dynamical variables over the Common Era. *Scientific Data*, *5*, 1–15. doi: 10.1038/sdata.2018.86

Sutton, R. T., McCarthy, G. D., Robson, J., Sinha, B., Archibald, A. T., & Gray, L. J. (2018, feb). Atlantic Multidecadal Variability and the U.K. ACSIS Program. *Bulletin of the American Meteorological Society*, *99*(2), 415–425. doi: 10.1175/BAMS-D-16-0266.1

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., . . . Noone, D. (2019, jul). Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. *Climate of the Past*, *15*(4), 1251–1273. doi: 10.5194/cp-15-1251-2019

Tardif, R., Hakim, G. J., & Snyder, C. (2014). Coupled atmosphere-ocean data assimilation experiments with a low-order climate model. *Climate Dynamics*, *43*(5-6), 1631–1643. doi: 10.1007/s00382-013-1989-0

Taylor, K. E., Stouffer, R. J., & Meehl, G. a. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. doi: 10.1175/BAMS-D-11-00094.1

Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., & Poulsen, C. J. (2020, aug). Glacial cooling and climate sensitivity revisited. *Nature*, *584*(7822), 569–573. doi: 10.1038/s41586-020-2617-x

Trigo, R. M., Osborn, T. J., & Corte-Real, J. M. (2002). The North Atlantic Oscillation influence on Europe: Climate impacts and associated physical mechanisms. *Climate Research*, *20*(1), 9–17. doi: 10.3354/cr020009

Trouet, V., Esper, J., Graham, N. E., Baker, A., Scourse, J. D., & Frank, D. C. (2009, apr). Persistent Positive North Atlantic Oscillation Mode Dominated the Medieval Climate Anomaly. *Science*, *324*(5923), 78–80. doi: 10.1126/science.1166349

Vecchi, G. A., Delworth, T. L., & Booth, B. (2017, aug). Origins of Atlantic decadal swings. *Nature*, *548*(7667), 284–285. doi: 10.1038/nature23538

Wallace, J. M., & Thompson, D. W. J. (1998). The Arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, *25*(9), 1297–1300.

Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P., & Rajaratnam, B. (2015). Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation. *Geophysical Research Letters*, *42*(17), 7162–7170. doi: 10.1002/2015GL065265

Whitaker, J., & Hamill, T. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 1913–1924.

Wills, R. C., Armour, K. C., Battisti, D. S., & Hartmann, D. L. (2019). Ocean-atmosphere dynamical coupling fundamental to the atlantic multidecadal oscillation. *Journal of Climate*, *32*(1), 251–272. doi: 10.1175/JCLI-D-18-0269.1

Wills, R. C., Schneider, T., Wallace, J. M., Battisti, D. S., & Hartmann, D. L. (2018). Disentangling Global Warming, Multidecadal Variability, and El Niño in Pacific Temperatures. *Geophysical Research Letters*, *45*(5), 2487–2496. doi: 10.1002/2017GL076327

Zanna, L. (2012). Forecast skill and predictability of observed atlantic sea surface temperatures. *Journal of Climate*, *25*(14), 5047–5056. doi: 10.1175/JCLI-D-11-00539.1

Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J., & Heimbach, P. (2019, jan). Global reconstruction of historical ocean heat storage and transport. *Proceedings of the National Academy of Sciences*, *116*(4), 1126–1131. doi: 10.1073/pnas.1808838115

Zhang, R., Sutton, R., Danabasoglu, G., Delworth, T. L., Kim, W. M., Robson, J., & Yeager, S. G. (2016, jun). Comment on "The Atlantic Multidecadal Oscillation without a role for ocean circulation". *Science*, *352*(6293), 1527–1527. doi: 10.1126/science.aaf1660

Supporting Information for "Coupled Atmosphere–Ocean Reconstruction of the Last Millennium Using Online Data Assimilation"

W. A. Perkins[1], G. J. Hakim[1]

[1]University of Washington

[1]Seattle, WA

Contents of this file

1. Text S1
2. Text S2
3. Figures S1 to S9
4. Tables S1 to S3

Additional Supporting Information (Files uploaded separately)

1. Captions for Movies S1 to S2

Introduction

The provided supporting information covers the process by which we determine the production LIM calibration parameters, and describes the details of sensitivity testing regarding the cooler temperatures near the beginning of the reconstruction (∼1000–1400

Corresponding author: W. A. Perkins, Department of Atmospheric Sciences, University of Washington (wperkins@uw.edu)

C.E.). Additionally, we provide figures and two movies as supporting information to the main text.

S1. LIM Calibration Testing

To test the ensemble forecast characteristics of the LIM relevant for data assimilation, we investigate ensemble calibration ratios (ECRs; e.g., as in Perkins & Hakim, 2020) for a number of multivariate-EOF (mvarEOF) component truncations (ℓ = 10, 15, 20, 21, 22, ..., 29, 30). For each test, we calibrate the LIM (Eq. 4) with the specified number of retained multivariate-EOF components and perform a 1-year ensemble forecast (100 ensemble members) initialized from every available year of the calibration data. ECRs are based on comparison to the reference calibration data coincident with the forecast time. The ECR measure, defined as

$$\mathrm{ECR} = \frac{1}{T}\sum_{i=1}^{T}\frac{\mathrm{SE}_t}{\sigma_t^2}, \qquad (1)$$

represents the time-average (over all times, T) ratio between squared errors (SE_t) and ensemble variance (σ_t^2) calculated on the forecast ensemble ($\mathbf{g}_t$) and reference data (v_t). The squared error at time t, is defined as

$$\mathrm{SE}_t = (\overline{\mathbf{g}}_t - v_t)^2,$$

where the overline (e.g., $\overline{\mathbf{g}}_t$) denotes the ensemble average. The ensemble variance is given by

$$\sigma_t^2 = \frac{1}{N}\sum_{i=1}^{N}(g_{ti} - \overline{\mathbf{g}})^2$$

where g_{ti} represents the i^{th} of N total ensemble members. Well calibrated ensemble forecasts have a value near 1.0, while values less than 1.0 are considered overdispersive (errors are smaller than ensemble spread) and values greater than 1.0 are considered underdispersive (errors are larger than ensemble spread).

After performing the ensemble forecast experiments, we compare ensemble characteristics between them by aggregating the relative distance from 1.0 for groups of ECR quantities (for global averages, ENSO, and PDO-related measures). The function, $f(\mathbf{g}, v)$ we use to calculate distance from being well-calibrated is as follows:

$$f(\mathbf{g}, v) = \begin{cases} ln(\mathrm{ECR}(\mathbf{g}, v)) & 0 < \mathrm{ECR}(\mathbf{g}, v) < 1 \\ (\mathrm{ECR}(\mathbf{g}, v) - 1)^2 & \mathrm{ECR}(\mathbf{g}, v) \geq 1 \end{cases}. \tag{2}$$

For each reconstruction experiment, we select the multivariate-EOF truncation that displays the lowest total ECR distance from 1.0 (Tables S2 and S3). For the CCSM4-LIM, the minimum aggregate ECR occurs at a truncation of 20 modes ($f = 0.79$), and for the MPI-LIM, the minimum occurs at 27 modes ($f = 1.25$).

S2. Pre-industrial Control LIM Test

Our reconstructed global-average temperature estimates during the early period of the last millennium are cooler than many previous reconstructions (Fig. 6). Here we investigate whether the cooler average temperature during the early period of the last millennium is caused by the LIM model formulation. Specifically, we test whether the inclusion of forcing (especially volcanic response) plays a role in the cool temperatures and lack of millennium-scale cooling trend, by training another LIM based on the CCSM4 pre-industrial control simulation (no-forcing) and using it for an online DA reconstruction.

To train the pre-industrial control (piControl) LIM, we use the same fields and procedures as in the past1000 CCSM4-LIM (see Sections 3 and S1), searching for a "well-calibrated" ensemble forecast. In general, the lower variance of the piControl simulation produces underdispersive forecast ensembles when testing forecasts against piControl data (minimum ECR at 26 modes retained, $f = 2.1$) and severely underdispersive results when testing forecasts against past1000 data (minimum at 22 modes retained, $f = 2550$). We select $l = 26$ for the number of retained multivariate-EOFs based on the forecast tests against piControl data. The $ECR >> 1$ values shown in the piControl calibration experiments indicate that variance inflation is necessary to produce a representative ensemble for data assimilation purposes. To find an appropriate inflation ratio, we perform three reconstructions (with 5 iterations each) with inflation factors of 1.0, 2.0, and 4.0, to find an inflation parameter that produces at least similar global-average TAS verification scores as in the offline case. The experiment with an inflation factor of 4.0 produced the closest verification correlations (SST ∼0.8, Niño 3.4 ∼0.7, PDO ∼0.6, OHC700m ∼0.1–0.5). Therefore, we use the inflation factor 4.0 experiment to assess whether the LIM calibrated on the past1000 data is solely responsible for the colder temperatures.

Results of the piControl-LIM experiment are compared against the offline case in Fig. S6. The global-average temperatures for the piControl-LIM (Fig. S6a) still show cooler conditions are prevalent during the early period (1000–1400 C.E.) with no clear millennium-scale cooling trend. The global-average SST and OHC700m also show cooler conditions compared to the offline case, albeit less consistently cool than in the CCSM4-LIM experiment from the main text. In general, the multi-decadal variations and uncertainty bounds are larger in the piControl-LIM experiment, which is related to our use of inflation. However, even without any information on forced-response of the climate system, the piControl-LIM reconstruction still produces colder temperatures. This strongly suggests the behavior during the early part of the reconstruction is a byproduct of the system memory introduced by the LIM and consistent reinforcement from proxy information during that time.

Movie S1.

A video showing the grand-ensemble mean (taken over the 50 × 100 Monte-Carlo iterations and ensemble members) spatial field results (TAS, SST, PR, SLP, ZG500) for the LMR Online reconstruction using the CCSM4-LIM calibration. Fields are centered about the 1000–1850 mean values. The spatial distribution of the proxy network available for assimilation in each year is provided in the lower right panel.

Movie S2.

As in Movie S1, but depicting fields of RLUT, RSUT, SST, OHC700m, SSS, and ZOS. Provided as a separate video for clarity.

References

Biondi, F., Gershunov, A., & Cayan, D. R. (2001). North Pacific Decadal Climate Variability since 1661. *Journal of Climate*, *14*(1), 5–10. doi: 10.1175/1520-0442(2001)014⟨0005:NPDCVS⟩2.0.CO;2

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., & Bladé, I. (1999). The effective number of spatial degrees of freedom of a time-varying field. *Journal of Climate*, *12*(7), 1990–2009. doi: 10.1175/1520-0442(1999)012⟨1990:TENOSD⟩2.0.CO;2

Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., & Zhu, J. (2017). Improved estimates of ocean heat content from 1960 to 2015. *Science Advances*, *3*(3), 1–11. doi: 10.1126/sciadv.1601545

D'Arrigo, R., Villalba, R., & Wiles, G. (2001). Tree-ring estimates of Pacific decadal climate variability. *Climate Dynamics*, *18*(3-4), 219–224. doi: 10.1007/s003820100177

D'Arrigo, R., & Wilson, R. (2006). On the Asian expression of the PDO. *International*

Journal of Climatology, *26*(12), 1607–1617. doi: 10.1002/joc.1326

MacDonald, G. M., & Case, R. A. (2005). Variations in the Pacific Decadal Oscillation over the past millennium. *Geophysical Research Letters*, *32*(8), 1–4. doi: 10.1029/2005GL022478

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. *Bulletin of the American Meteorological Society*, *78*(6), 1069–1079. doi: 10.1175/1520-0477(1997)078⟨1069:APICOW⟩2.0.CO;2

Perkins, W. A., & Hakim, G. (2020). Linear inverse modeling for coupled atmosphere-ocean ensemble climate prediction. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001778. (e2019MS001778 10.1029/2019MS001778) doi: 10.1029/2019MS001778

Shen, C., Wang, W. C., Gong, W., & Hao, Z. (2006). A Pacific Decadal Oscillation record since 1470 AD reconstructed from proxy data of summer rainfall over eastern China. *Geophysical Research Letters*, *33*(3), 1–4. doi: 10.1029/2005GL024804

Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J., & Heimbach, P. (2019, jan). Global reconstruction of historical ocean heat storage and transport. *Proceedings of the National Academy of Sciences*, *116*(4), 1126–1131. doi: 10.1073/pnas.1808838115

Table S1. LMR Online (MPI-LIM) reconstruction scalar correlations with instrumental products.

Product	Glob. Avg. SST	Glob. Avg. OHC700m	Nino 3.4	PDO
HadleyEN4	0.79	0.37	0.72	0.42
GFDLECDA	0.88	0.56	0.78	0.24
ORA20C	0.91	0.85	0.72	0.55
ESRL	–	–	0.77	–
Mantua	–	–	–	0.42
Cheng17	–	0.90	–	–
Zanna19	–	0.98	–	–

Table S2. Aggregate ensemble calibration ratio (ECR) distances from the ideal ECR of 1.0 as calculated using Eq. S2. Values are shown for each test using different mvarEOF truncations during LIM calibration on CCSM4 last millennium data. The global average (Glob Avg) measure includes ECRs from TAS, SST, OHC700m, RSUT, and RLUT. The ENSO measure includes ECRs from Niño 3, 3.4, and 4 indices and the Southern Oscillation Index (SOI). The PDO measure only includes the PDO index ECR. The Total column shows the sum of Glob Avg, ENSO, and PDO aggregate ECR distances.

num mvarEOFs	Glob Avg	ENSO	PDO	**Total**
15	2.98	0.20	0.01	3.18
20	0.14	0.00	0.65	0.79
21	0.42	0.01	0.71	1.15
22	0.21	0.01	0.88	1.11
23	0.32	0.03	0.82	1.17
24	1.75	0.26	0.21	2.22
25	0.01	0.51	2.70	3.22
26	3.27	0.24	0.03	3.55
27	0.54	0.08	0.46	1.09
28	0.84	0.22	0.12	1.17
29	1.33	0.11	0.52	1.96
30	1.60	0.03	0.09	1.73

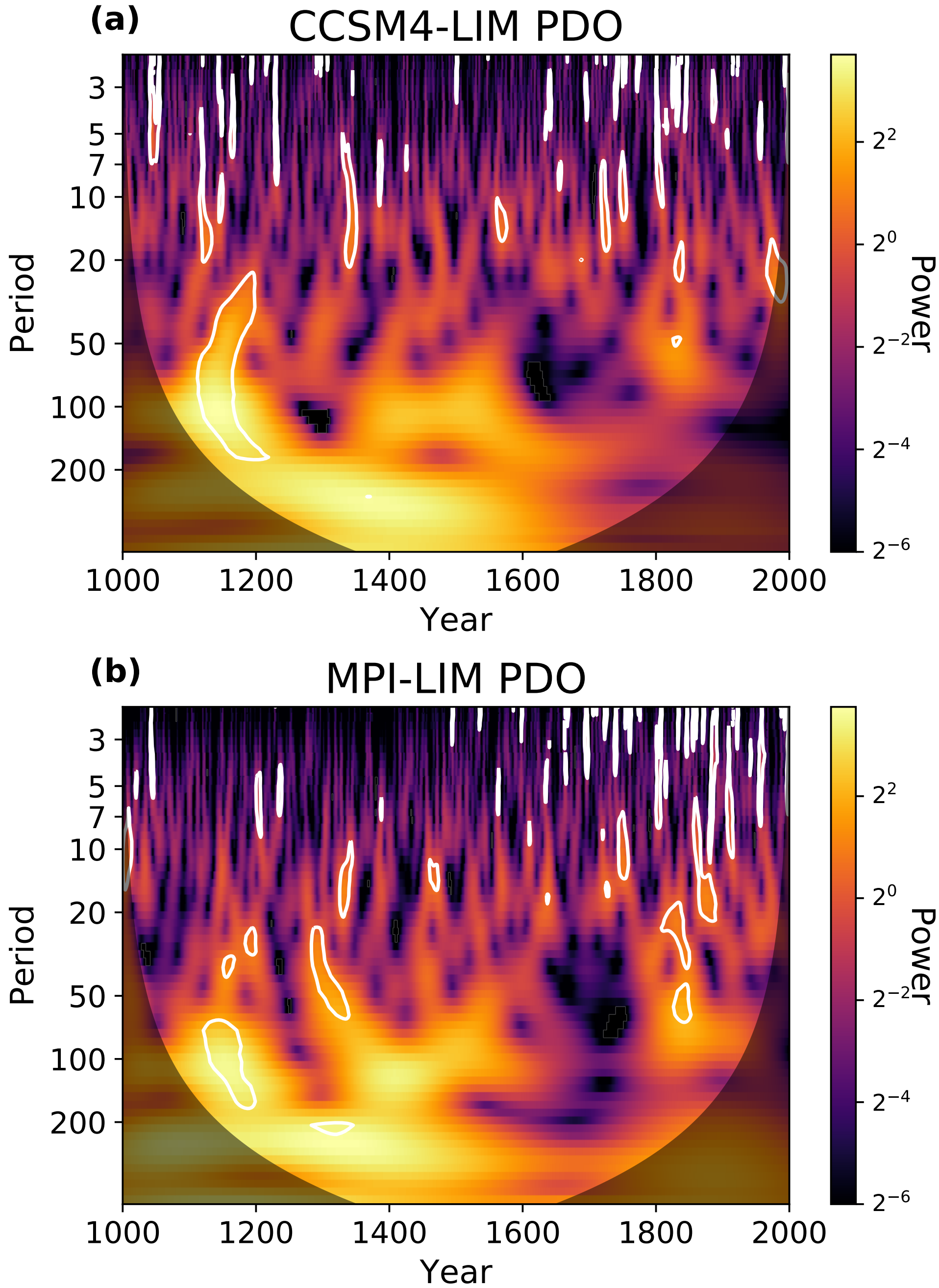

Figure S1. Complex Morlet wavelet analysis of the grand-ensemble average (taken over 50 × 100 members) Pacific Decadal Oscillation (PDO) index reconstructed in the (a) CCSM4-LIM and (b) MPI-LIM experiments. The displayed power is normalized by dividing by the input timeseries variance in each case. White contours indicate a power exceedance of a 95% confidence interval generated using 1000 integrations of a red-noise model fit to the PDO timeseries.

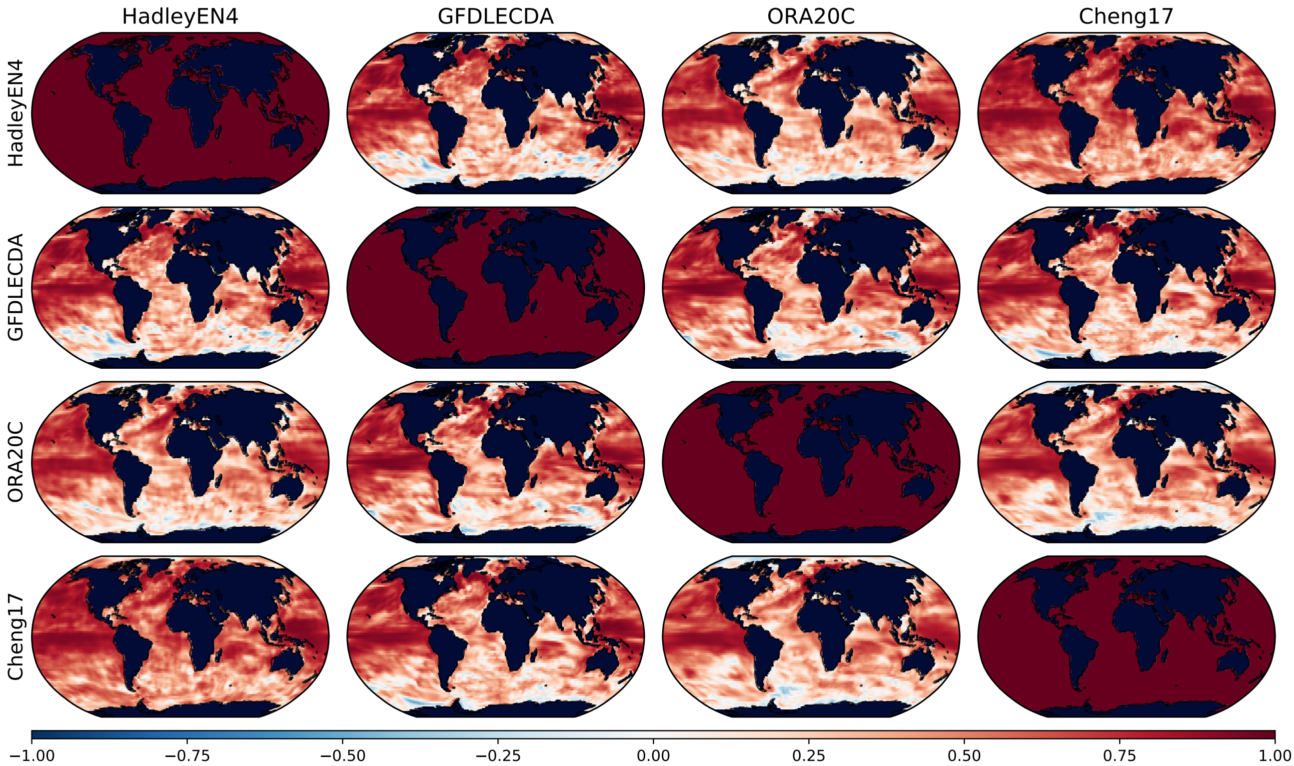

Figure S2. A gridpoint correlation comparison of the OHC700m spatial validation products with each other including HadleyEN4 (1900–2000), GFDLECDA (1961–2000), ORA-20C (1900–2000), and Cheng2017 (1940–2000)

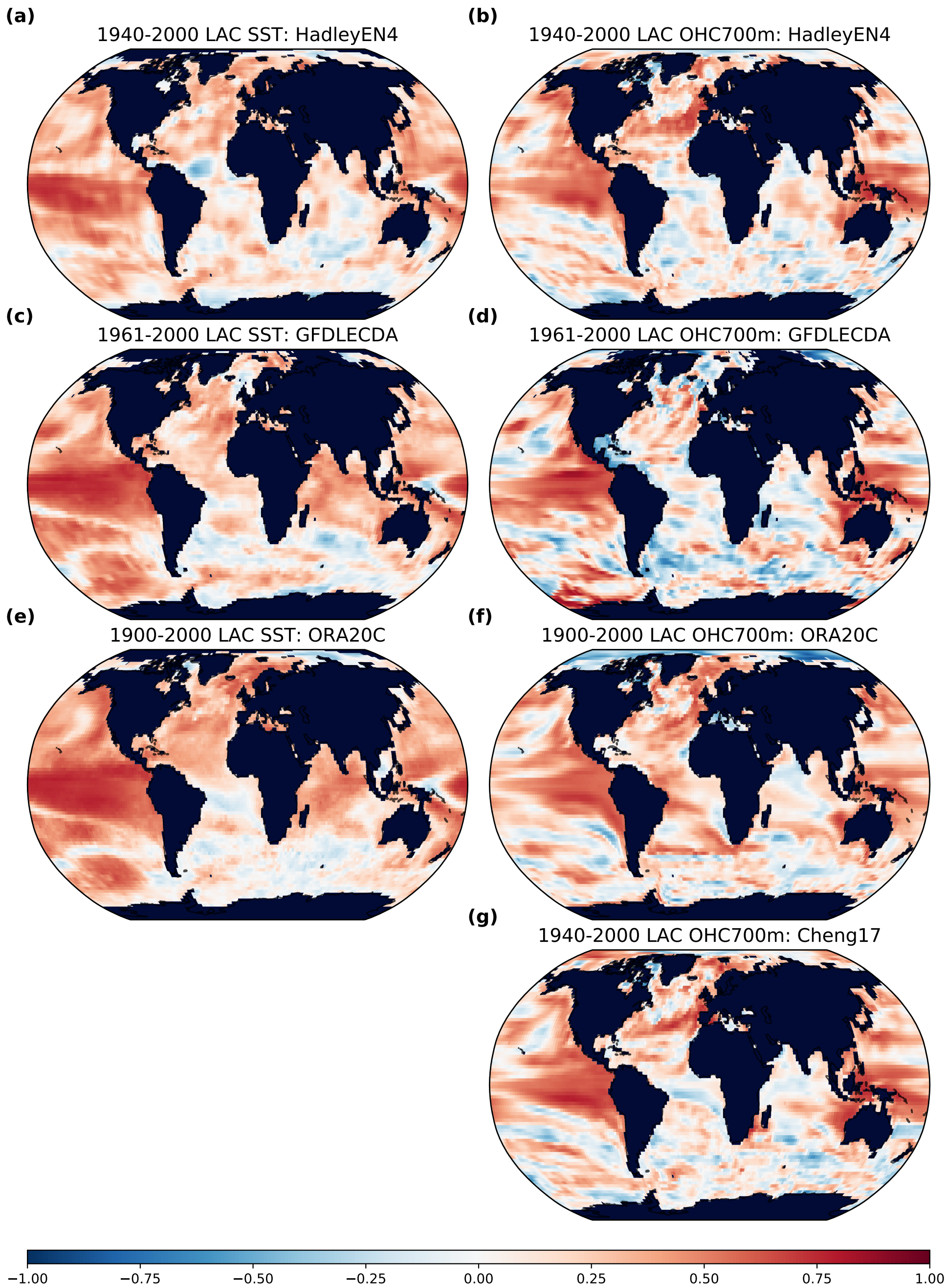

Figure S3. Detrended spatial field gridpoint correlations of the LMR MPI-LIM with Instrumental Era observational and reanalysis products for SSTs (column a) and OHC700m (column b). Spatial correlations are calculated against HadleyEN4 data (a, b; 1950-2000), GFDLECDA (c, d; 1961–2000), ORA-20C (e, f; 1900–2000), and Cheng2017 (OHC only, g; 1940–2000)

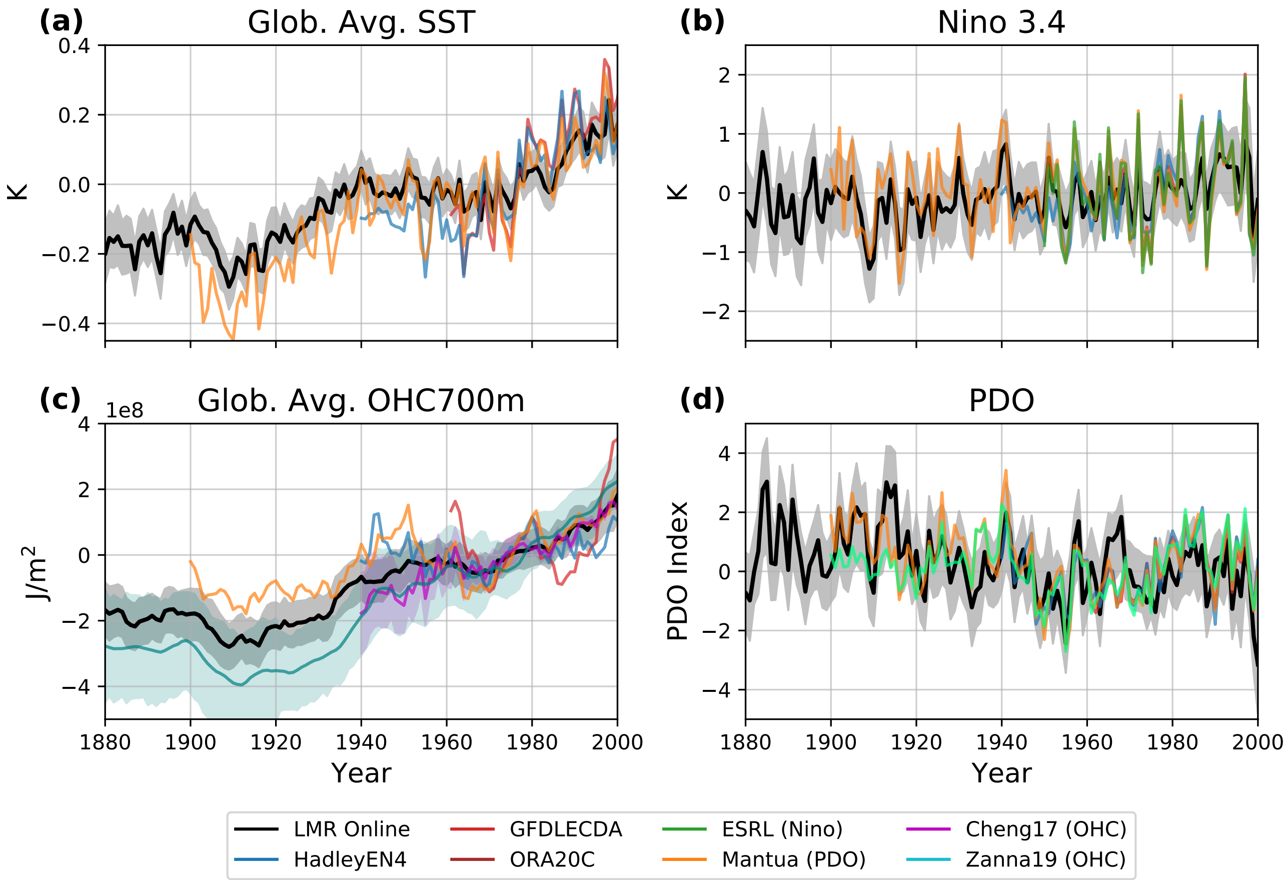

Figure S4. Scalar index comparison between the LMR Online (MPI-LIM; black with 95% confidence bounds in grey shading) reconstruction and instrumental products for (a) SST, (b) Niño 3.4, (c) OHC700m, and (d) PDO. The HadleyEN4, GFDLECDA, and ORA20C products are compared in all cases. Additionally, ESRL Niño 3.4 data, the Mantua et al. (1997) PDO index, and Cheng et al. (2017, Cheng2017) and Zanna et al. (2019, Zanna19) OHC data are compared. Error bounds ($\pm 2\sigma$) are shown for the Cheng2017 and Zanna19 OHC data.

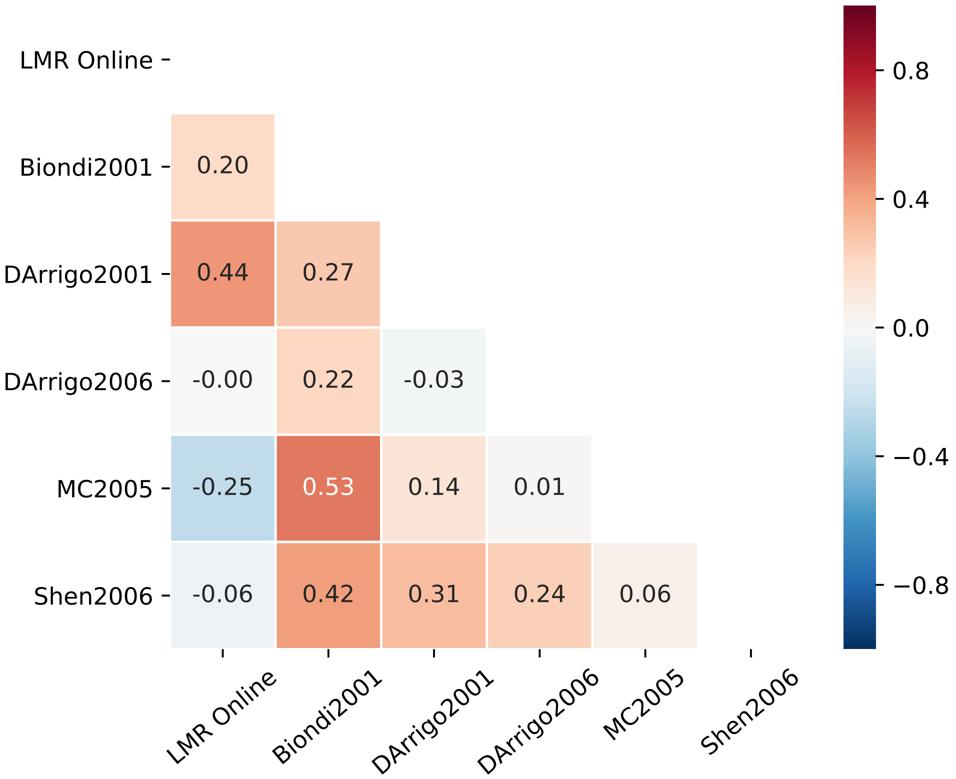

Figure S5. Correlations between the PDO index of the CCSM4-LIM reconstruction and previous PDO index reconstructions (Biondi et al., 2001; D'Arrigo et al., 2001; MacDonald & Case, 2005; D'Arrigo & Wilson, 2006; Shen et al., 2006).

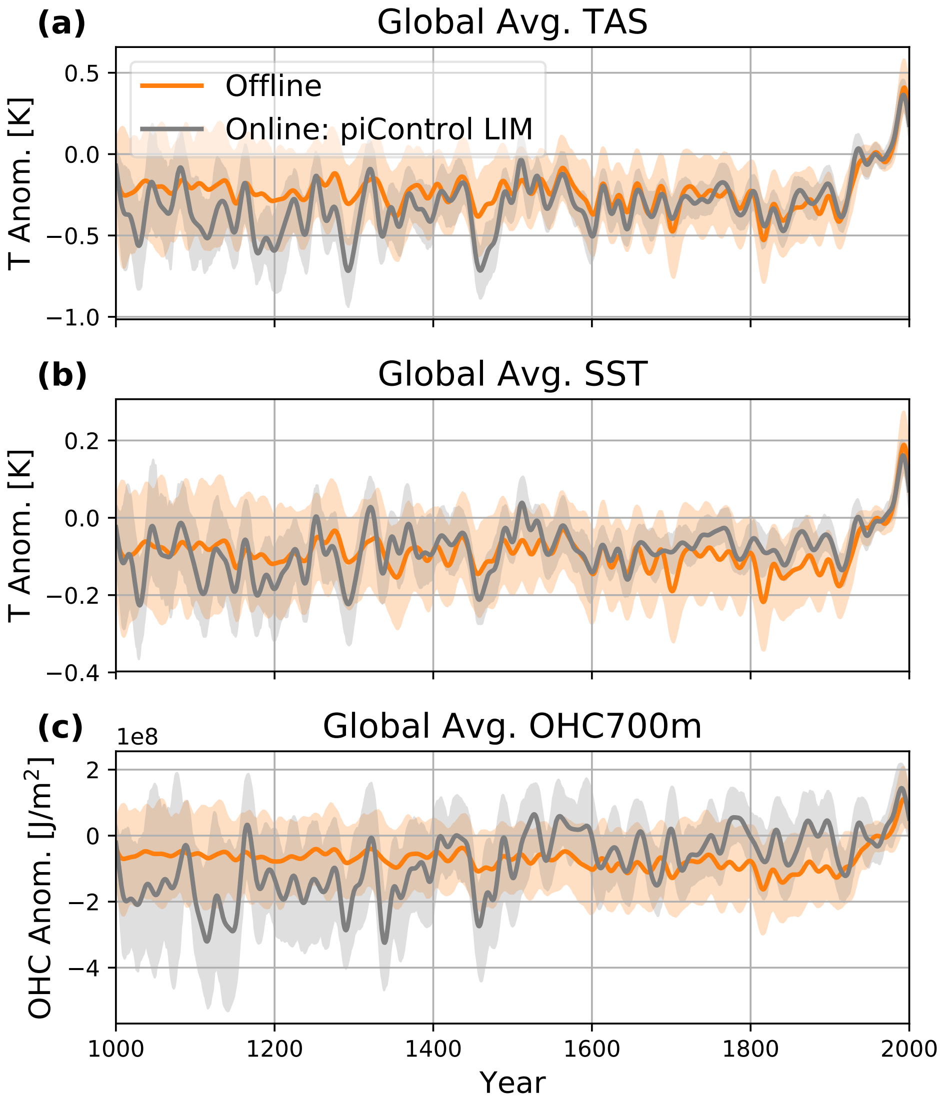

Figure S6. As in Fig. 7, but comparing the offline reconstruction against an online reconstruction using a LIM calibrated against CCSM4 preindustrial-control data. See Section S2 for details.

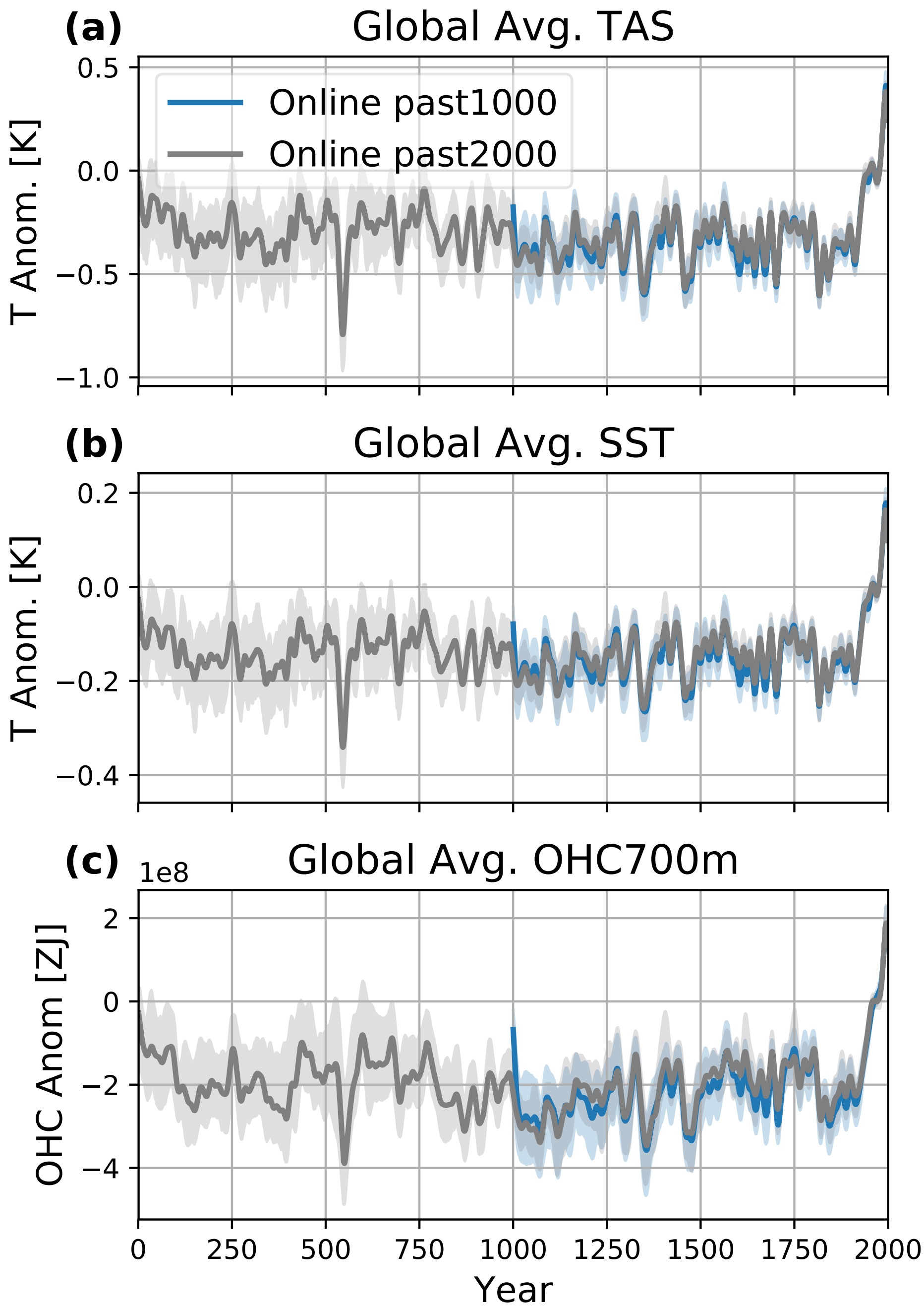

Figure S7. As in Fig. 7, but comparing online reconstructions intialized at 1000 C.E. (past1000) and 1 C.E. (past2000).

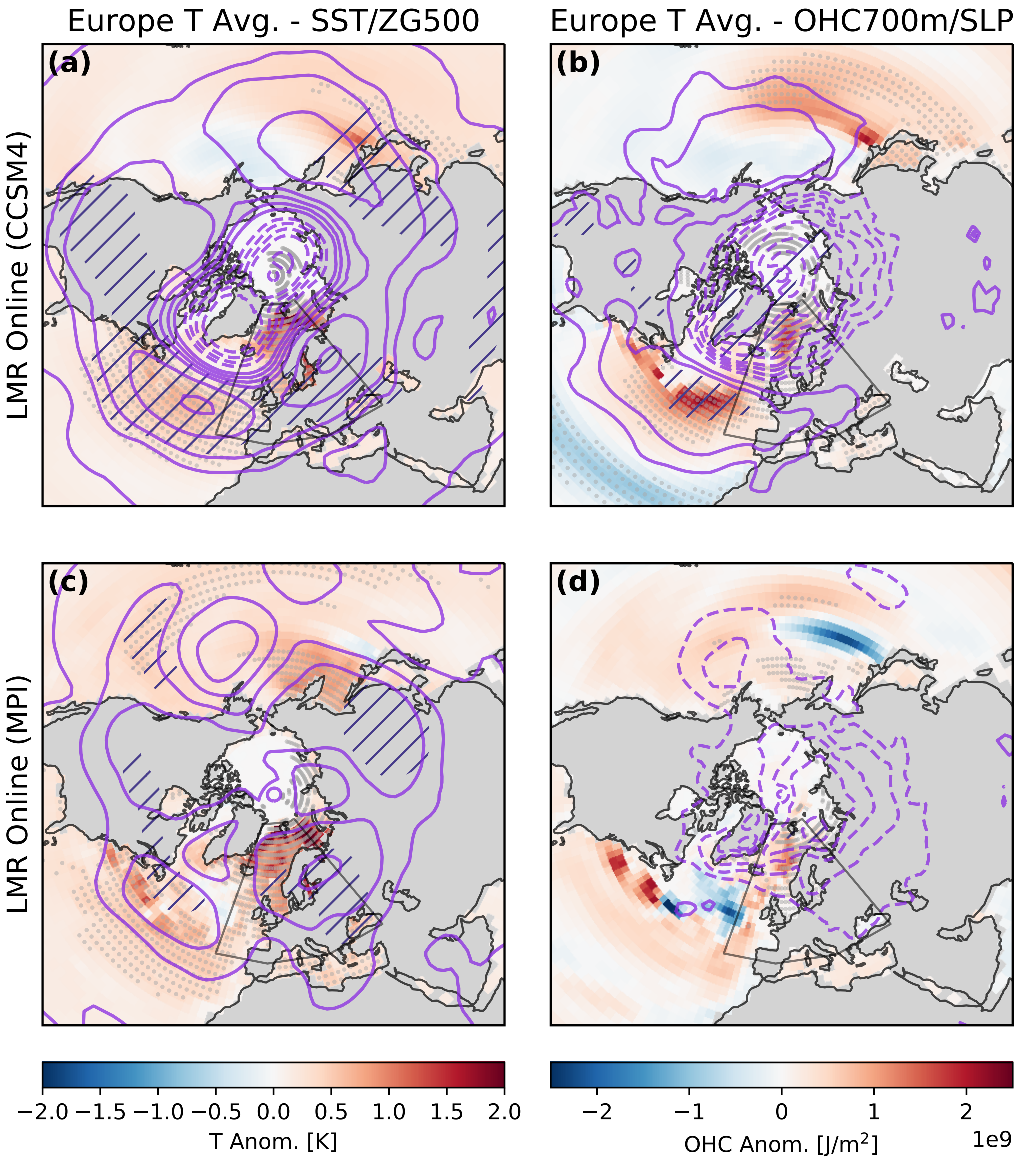

Figure S8. Regression of the reconstructed Europe average TAS (region denoted by black box) from 1000–1850 C.E. onto fields of SST/ZG500 (column a) and OHC700m/SLP (column b) for the CCSM4-LIM (row a) and MPI-LIM (row b) reconstructions. ZG500 field contour levels range from 7.5 m to 20 m incremented every 2.5 m for positive (solid) and negative (dashed) values. SLP field contour levels range from 0.25 hPa to 1.25 hPa incremented every 0.25 hPa. Regression coefficient significance (grey dots for SST/OHC700m, blue hatching for ZG500/SLP) determined using a two-tailed Student's t-test and the effective degrees of freedom (Bretherton et al., 1999).

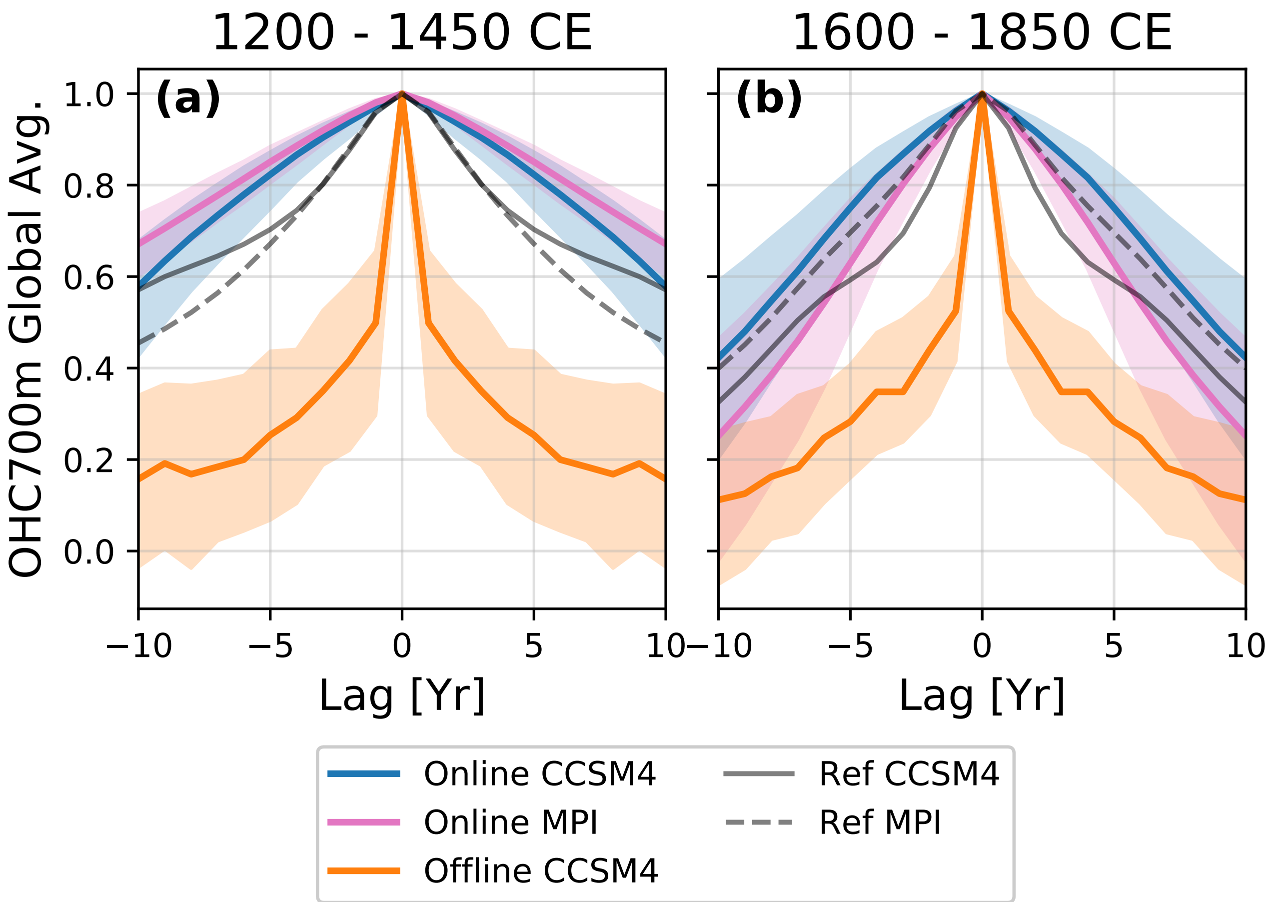

Figure S9. Autocorrelations of global average OHC700m for the time periods of 1200–1450 C.E. (a) and 1600–1850 C.E (b). Autocorrelations are shown for the Online CCSM4-LIM (blue) and MPI-LIM (pink) reconstructions, the CCSM4 Offline (orange) reconstruction, and reference Last Millennium Simulation data from the CCSM4 (black solid) and MPI (dashed black) models. Correlations are calculated at the specified lead/lag from the ensemble mean of each of the 50 reconstruction Monte-Carlo iterations. For reconstructions, solid lines denote the average correlation across Monte-Carlo iterations while shading shows the 95% confidence interval.

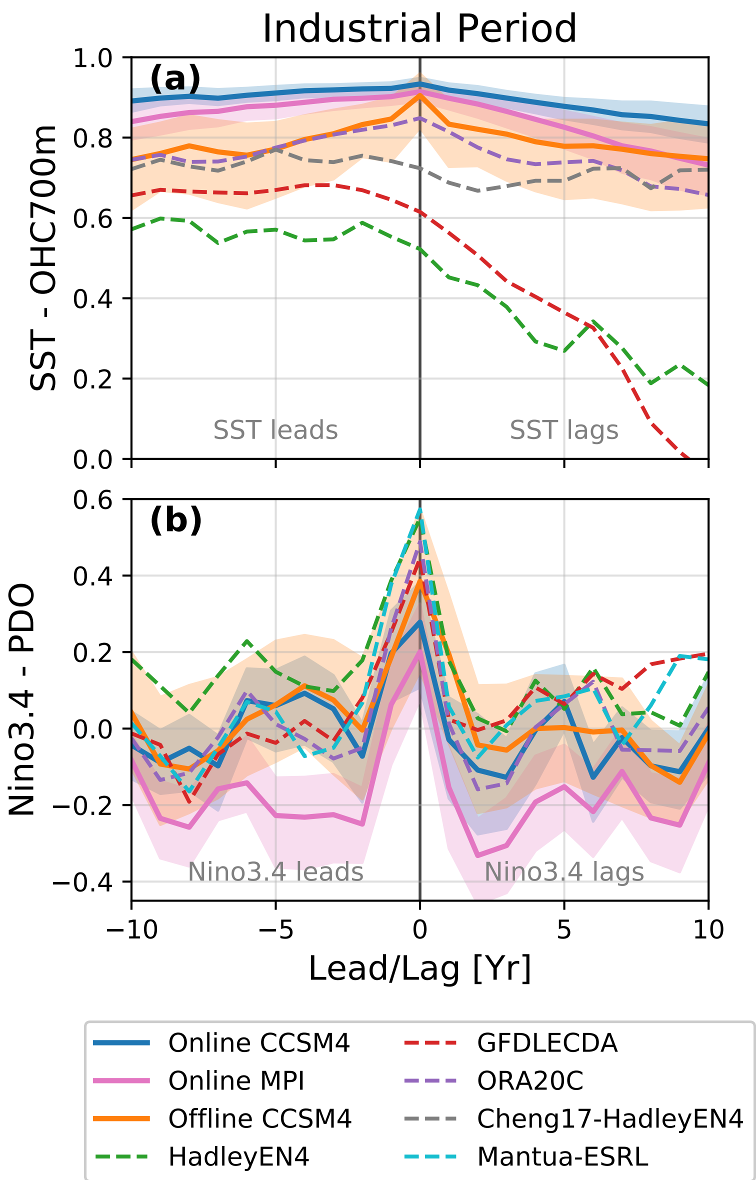

Figure S10. Lead–lag correlations for scalar indices of global average SST and OHC700m (a) and Niño 3.4 and PDO (c) during the instrumental period from 1900–2000 C.E. Correlations are shown for the Online CCSM4-LIM (blue) and MPI-LIM (pink) reconstructions, the CCSM4 Offline (orange) reconstruction, and instrumental products (dashed). Correlations are calculated at the specified lead/lag from the ensemble mean of each of the 50 reconstruction Monte-Carlo iterations. For reconstructions, solid lines denote the average correlation across Monte-Carlo iterations while shading shows the 95% confidence interval.

Table S3. As in Table S2, but for LIMs calibrated on MPI last millennium data. The global average (Glob Avg) measure includes ECRs from TAS, SST, OHC700m. RSUT and RLUT were omitted due to large ECR values dominating the selection process and the focus on atmosphere–ocean data in the present study. The ENSO measure includes ECRs from Niño 3, 3.4, and 4 indices and the Southern Oscillation Index (SOI). The PDO measure only includes the PDO index ECR. The Total column shows the sum of Glob Avg, ENSO, and PDO aggregate ECR distances.

num mvarEOFs	Glob Avg	ENSO	PDO	**Total**
15	1.07	0.12	0.51	1.69
20	3.14	0.13	0.18	3.45
21	1.96	0.35	0.09	2.40
22	1.47	0.11	0.18	1.76
23	1.95	0.02	0.53	2.50
24	0.25	0.43	0.75	1.42
25	2.24	0.14	0.26	2.64
26	1.49	0.15	0.11	1.75
27	1.03	0.20	0.02	1.25
28	1.56	0.05	0.82	2.43
29	1.94	0.28	0.12	2.34
30	5.22	0.84	0.18	6.24