Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

Climate of the Past
Discussions

Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling

Robert Tardif[1], Gregory J. Hakim[1], Walter A. Perkins[1], Kaleb A. Horlick[2], Michael P. Erb[3], Julien Emile-Geay[4], David M. Anderson[5], Eric J. Steig[6,1], and David Noone[2]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA
[2]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA
[3]School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, AZ, USA
[4]Department of Earth Sciences, University of Southern California, Los Angeles, California, USA
[5]Retired, NOAA Paleoclimatology Program, Boulder, CO, USA
[6]Department of Earth and Space Sciences, University of Washington, Seattle, Washington, USA

Correspondence: R. Tardif (rtardif@atmos.washington.edu)

Abstract.

The Last Millennium Reanalysis utilizes an ensemble methodology to assimilate paleoclimate data for the production of annually resolved climate field reconstructions of the Common Era. Two key elements are the focus of this work: the set of assimilated proxy records, and the forward models that map climate variables to proxy measurements. Results based on an extensive proxy database and seasonal regression-based forward models are compared to the prototype reanalysis of Hakim et al. (2016), which was based on a smaller set of proxy records and simpler proxy models formulated as univariate linear regressions against annual temperature. Validation against various instrumental–era gridded analyses shows that the new reconstructions of surface air temperature, 500 hPa geopotential height and the Palmer Drought Severity Index are significantly improved, with skill scores increasing from 10% to more than 200%, depending on the variable and verification measure. Additional experiments designed to isolate the sources of improvement reveal the importance of additional proxy records, including coral records for improving tropical reconstructions; tree-ring-width chronologies, including moisture-sensitive trees, for thermodynamic and hydroclimate variables in mid-latitudes; and tree-ring density records for temperature reconstructions, particularly in high northern latitudes. Proxy forward models that account for seasonal responses, and the dual sensitivity to temperature and moisture characterizing tree-ring-width proxies, are also found to be particularly important. Other experiments highlight the beneficial role of covariance localization on reanalysis ensemble characteristics. This improved paleoclimate data assimilation system served as the basis for the production of the first publicly released NOAA Last Millennium Reanalysis.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

1 Introduction

Reconstructions of Earth's past climate, particularly covering periods prior to instrumental data sets, are key to understanding the causes of natural climate variability. For example, understanding natural variability provides the basis for improving predictions of climate variability in the coming decades. Information on past climates has traditionally been derived either from climate proxy data (e.g., tree rings, ice cores, etc.) or from Earth system model simulations. Synthesizing information from these two sources is a central challenge of paleoclimate science. Paleoclimate data assimilation (PDA) has emerged as a powerful framework for such synthesis because it provides the optimal combination of climate signals recorded by proxies as constrained by the dynamics of Earth system models. PDA-generated climate field reconstructions have been used to investigate climate variability prior to the instrumental era (Goosse et al., 2006, 2010; Widmann et al., 2010; Bhend et al., 2012; Steiger et al., 2014; Matsikaris et al., 2015; Franke et al., 2017; Okazaki and Yoshimura, 2017; Steiger et al., 2018). Within this general PDA framework, a flexible PDA system is being developed for the Last Millennium climate Reanalysis (LMR) project for the production of annually resolved reconstructions of the Common Era. Hakim et al. (2016) describe a prototype configuration of the LMR and show results in good agreement with previous reconstructions of Northern Hemisphere mean near-surface air temperature. Detailed comparisons with several gridded instrumental temperature data products revealed significant skill over tropical regions but less skillful reconstructions over Northern Hemisphere continental areas, where a large proportion of proxy data are located.

As with any data assimilation system, two important components impacting the quality of the resulting analyses are the set of assimilated observations (here, proxy records) and the forward models that map variables from climate model output to proxy measurements ("proxy system models"; hereafter, PSMs). Hakim et al. (2016) assimilated proxy records from the first compilation of the PAGES 2k Consortium (PAGES 2k Consortium, 2013), and modeled the proxies through univariate linear regressions calibrated against annually-averaged instrumental temperature data. Here we examine the impact on LMR reconstructions of improvements to these two key components: (1) a much larger proxy data base, with records from PAGES 2k Consortium (2017), Breitenmoser et al. (2014) and additional records described in Anderson et al. (2018); (2) more realistic PSMs in which seasonality and, for tree-ring-width proxies, temperature and moisture sensitivity are taken into account. Motivation for expanding the proxy database derives from evidence that climate reconstructions are generally sensitive to the set of proxy records used as input (e.g., Wang et al., 2015), while the introduction of more sophisticated PSMs is in part motivated by the fact that comprehensive reconstructions of temperature and hydroclimate variables depend on properly treating temperature-sensitive and moisture-sensitive tree ring proxies (e.g., Steiger and Smerdon, 2017).

The focus of improvements in PSMs here is on regression-based (i.e. statistical) forward models, in contrast to recent efforts focusing on process-based PSMs (see e.g., Breitenmoser et al., 2014; Dee et al., 2016; Acevedo et al., 2017). Our objective is to establish baseline skill of PDA reconstructions using statistical PSMs, to serve as a benchmark for evaluating possible improvements associated with process-based PSMs (e.g., Dee et al., 2016). Here we develop a hierarchy of statistical PSMs to identify aspects that contribute increased skill to reconstructed temperature and hydroclimate states.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

The paper is organized as follows. Section 2 outlines the LMR PDA-based framework and describes the proxy database and PSMs. Reconstructions based on this configuration are presented in section 3, with comparisons to the prototype described in Hakim et al. (2016). Section 4 explores the contributions to improvements in the new reconstructions. A concluding summary is given in section 5.

2 Methods

Paleoclimate data assimilation has three main components: proxy records, providing an indirect record of past climatic conditions; climate models, providing prior estimates of the climate; and proxy system models, providing the connection between the model prior and the proxy values. The method for each component is now described.

2.1 Data assimilation framework

LMR employs ensemble data assimilation (DA) to optimally blend information from proxies and climate model data. DA is performed using a variant of the ensemble Kalman filter, which for our application appears to perform well compared to alternative PDA methods such as particle filters (Liu et al., 2017). The update equation is given by

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y} - \mathbf{y}_e]. \tag{1}$$

Here, $\mathbf{x}_b$ is the prior state vector, which contains the climate variables to be reconstructed, averaged over an appropriate timescale (here, annual), while $\mathbf{x}_a$ is the posterior state vector (i.e. the reanalysis, or reconstruction). The state vector may include scalars, such as climate indices, and/or grid-point data for spatial fields. Vector $\mathbf{y}$ contains the assimilated proxy data (i.e. observations) and $\mathbf{y}_e$ is a vector containing estimates of the proxies derived from the prior by

$$\mathbf{y}_e = \mathcal{H}(\mathbf{x}_b), \tag{2}$$

where $\mathcal{H}$ is the forward model mapping the prior $\mathbf{x}_b$ to proxy space (i.e. the PSM, see section 2.4). The innovation, $[\mathbf{y} - \mathbf{y}_e]$, is the new information from the proxies not already contained in the prior. This new information is weighted against the prior through the Kalman gain matrix

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}\left[\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right]^{-1} \tag{3}$$

where $\mathbf{B}$ is the prior covariance matrix, $\mathbf{R}$ is the error covariance matrix for the proxy data, and $\mathbf{H}$ is the linearization of $\mathcal{H}$ about the mean value of the prior. Here, Eq. (1) is solved using the ensemble square-root filter (EnSRF) approach of Whitaker and Hamill (2002), in which the ensemble-mean and perturbations about the ensemble mean are solved separately. Moreover, $\mathbf{R}$ is taken as a diagonal matrix (uncorrelated observation errors) where the diagonal elements represent the error variance for each assimilated proxy record; details on how these are estimated are provided in Section 2.4. This allows for serial processing of observations, in which observations are assimilated one at a time. This greatly simplifies the implementation of

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

covariance localization, which is used to control sampling error in the prior covariance. Solutions for the ensemble mean, $\overline{\mathbf{x}}_a$, and perturbations, $\mathbf{x}'_a$, for the single k^{th} proxy y_k, are obtained from:

$$\overline{\mathbf{x}}_a = \overline{\mathbf{x}}_b + \frac{w_{loc} \circ cov(\mathbf{x}_b, y_{e,k})}{var(y_{e,k}) + R_k} (y_k - \overline{y}_{e,k}) \tag{4a}$$

$$\mathbf{x}'_a = \mathbf{x}'_b - \left[1 + \sqrt{\frac{R_k}{var(y_{e,k}) + R_k}}\right]^{-1} \frac{w_{loc} \circ cov(\mathbf{x}_b, y_{e,k})}{var(y_{e,k}) + R_k} (y'_{e,k}) \tag{4b}$$

where $y_{e,k}$ is the prior estimate of the proxy from (2) and R_k is the diagonal element of $\mathbf{R}$ corresponding to proxy y_k. The ensemble of updated states is then recovered by combining the posterior ensemble mean and perturbations

$$\mathbf{x}_a = \overline{\mathbf{x}}_a + \mathbf{x}'_a. \tag{5}$$

Covariance localization, given by a Schur product denoted by $\circ$ in Eqs. 4 (i.e. element-wise multiplication), is a distance-weighted filter w_{loc} on the prior covariance matrix (see e.g. Hamill et al., 2001). Section 4.3 provides for more details on localization. We also use an "appended state vector" approach for serially processing proxy measurements that avoids the need to recompute (2) after each proxy is assimilated. This approach also simplifies the implementation of seasonal forward models as described in more detail in Appendix A.

As in Hakim et al. (2016), an "offline" DA approach is used, where the prior ensemble is formed by random draws of time–averaged states from a pre-existing millennial-long model simulation, with the same randomly drawn ensemble members used for every year in the reconstruction. This is in contrast to online DA (e.g., Matsikaris et al., 2015; Perkins and Hakim, 2017), where a numerical model is used to dynamically forecast the evolution of climate states from the latest proxy-informed analysis to the following year, when new proxy observations are assimilated. The "offline" approach offers several practical advantages, particularly from a computational cost perspective. Furthermore, it is justified when model forecasts have little skill over timescales corresponding to the time interval between updates, as is the case here with proxies assimilated on an annual basis.

2.2 Climate proxies

Here we use a combination of proxy records from the latest PAGES 2k collection (PAGES 2k Consortium, 2017) and the additional records assembled by Anderson et al. (2018)[1]. As in the LMR prototype (Hakim et al., 2016, hereafter H16), only records with sub-annual to annual resolutions are considered; sub-annual records are averaged to annual. The PAGES 2k Consortium (2017) dataset (hereafter PAGES2k-2017) is a community standard in global proxy observations covering the Common Era (CE), and serves as the core source of proxy information used here. Proxies included in PAGES2k-2017 were screened to retain temperature-sensitive records only, extensively quality controlled, and described by more metadata compared to previous collections. Figure 1 shows the various proxy networks considered for assimilation, comparing the PAGES 2k Consortium (2013) (hereafter PAGES2k-2013) dataset used in H16, to the PAGES2k-2017 update, and to the entire updated

[1]An exception is the use of the Palmyra coral record from Cobb et al. (2003) rather than the Emile-Geay et al. (2013) update, as described in Anderson et al. (2018).

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

LMR database including the records of Anderson et al. (2018). Only records for which a PSM can be established are shown, defined by proxy records with at least 25 years of (non-contiguous) overlap with calibration data (see section 2.4).

Compared to the proxies assimilated in H16, PAGES2k-2017 data provide enhanced spatial coverage in the tropics with additional coral δ^{18}O and Sr/Ca records. Additional tree ring wood density records from Europe and western North America are also included. The temporal distribution of the total number of records remains similar, except for significant increases in the number of tree ring width and coral proxies during 1800–2000CE, and tree ring wood density records during 1500–2000CE. The additional records from Anderson et al. (2018) include the large number of tree ring width records from Breitenmoser et al. (2014) (hereafter B14), not strictly screened for temperature sensitivity in contrast to the PAGES 2k collection. These records provide enhanced coverage over eastern North America, southern Europe, boreal Eurasia and southern South America. Two tree ring records are also added in the data-void region of southern Africa. Other additions, totaling 94 records, provide additional records in the Tropics (23 coral records), and a greatly enhanced number of ice core records concentrated over Greenland and eastern Canadian Arctic (37 records) and Antarctica (26 records in West Antarctica and Drönning Maud Land). A few lower latitude ice core records (6 records) are also added in the Peruvian Andes and Tibetan Plateau, along with two higher latitude lake core records. From a temporal perspective relative to proxies used in H16, the addition of the tree ring width records from B14 contributes a notable number of additional proxies back to 1000CE, more than double the number of records available for assimilation from 1500CE onward, up to a fourfold increase during the 19th and 20th centuries.

2.3 Climate model prior information

For all reconstruction experiments reported in this paper, the prior state vector is formed with data from the CMIP5 (Taylor et al., 2012) Last Millennium simulation from the Community Climate System Model version 4 (CCSM4) coupled atmosphere-ocean-sea ice model. The simulation covers years 850 to 1850 CE and includes incoming solar variability, variable greenhouse gases as well as stratospheric aerosols from volcanic eruptions known to have occurred during the simulation period (see Landrum et al., 2013). The same "offline" DA methodology as in H16 is used, where the prior ensemble is a random sample of annual averages, with the same sample used for all years of the reconstruction. The sampled states are deviations (i.e. anomalies) from the temporal mean taken over the entire length of the simulation. Therefore, the prior ensemble–mean does not contain time-specific information about climate events (e.g. a volcanic eruption) or trends characterizing specific periods (e.g. twentieth century warming). Finally, the spatial resolution of prior state variables is reduced from $0.95^o \times 1.25^o$ of the Last Millennium simulation to a of $4.3^o \times 5.7^o$ Gaussian grid.

All reconstruction experiments are composed of 51 Monte-Carlo assimilation realizations, each using a different randomly chosen 100–member ensemble and 75% of available proxy records for assimilation. This Monte-Carlo sampling over subsets of prior states and proxy records is designed to incorporate uncertainties in covariance estimates derived from model states, and uncertainties associated with proxy error estimates. In the following, climate reanalyses are taken as the mean over the 100–member DA ensembles and 51 Monte-Carlo realizations (i.e., a 5100–member "grand ensemble").

2.4 Proxy modeling

A critical component of PDA is the mapping of prior climate state variables (e.g., temperature, precipitation from a climate model) to the assimilated proxies (e.g., tree ring width). This is expressed mathematically by Eq. 2, section 2.1, where the operator $\mathcal{H}$ (i.e. the forward model) ideally represents the complete set of processes associated with proxy values, i.e. a comprehensive physically-based PSM. This remains a major challenge as the information archive is often complex, involving physical, biological and chemical processes (Evans et al., 2013). Despite recent progress in the development and use of process-based PSMs (e.g., Dee et al., 2015, 2016; Goosse, 2016; Steiger et al., 2017; Acevedo et al., 2017), the focus here is on statistical PSMs, which offer distinct advantages: 1) ease of implementation and flexibility with respect to forward modeling of multiple proxies, regardless of archive types, measurements, units etc.; 2) observation error statistics for each assimilated record are well-defined from the regression (see below); and 3) regressions are formulated on the basis of deviations from the mean over a reference period (e.g. 1951–1980) of the driving climate variable(s), therefore avoiding issues with absolute calibration where climate model bias is problematic, particularly for PSMs having threshold transitions (see e.g., Dee et al., 2016). Statistical PSMs also have distinct disadvantages: 1) PSMs cannot be calibrated without sufficient overlap with calibration data; 2) the accuracy of the models depends on the limitations of the calibration datasets (e.g. less reliable analysis over the Southern Ocean and over high latitude continental areas due to a lack of observations); and 3) possible lack of stationarity of the derived relationships established with instrumental–era data. Despite these limitations, we believe statistical PSMs provide advantageous capabilities within the context of the LMR and, moreover, define a baseline to measure future progress with the development of process-based PSMs.

Here, univariate and bivariate statistical PSMs are considered,

$$y_k = \beta_{0k} + \beta_{1k}\overline{X_1'} + \epsilon_k \tag{6}$$

and

$$y_k = \beta_{0k} + \beta_{1k}\overline{X_1'} + \beta_{2k}\overline{X_2'} + \epsilon_k \tag{7}$$

where y_k are annualized observations from the k^{th} proxy time series, X_1', X_2' are anomalies of key climate variables (e.g. near-surface air temperature and precipitation) from calibration instrumental–era datasets, β_0 is the intercept and β_1, β_2 are the slopes with respect to the X_1' and X_2' independent variables respectively, and ϵ is a Gaussian random variable with zero mean and variance σ^2. The overbar in Eqs. 6 and 7 denotes time averages over annual periods as in H16, or over appropriate seasonal intervals for the seasonal PSMs. Calibration data concurrent with available proxy observations are taken at the grid point nearest the proxy location and the appropriate least-squares solution determines regression parameters $(\beta_0, \beta_1, \beta_2, \sigma)$. In this version of LMR, PSM configuration is the same for each proxy category (e.g., univariate for all coral $\delta^{18}O$, bivariate for all tree ring widths records, etc.).

With the framework described above, the regression–based approach measures the diagonal elements in matrix $\mathbf{R}$ through the variance of regression residuals, i.e. $R_k = \sigma^2$. This is a key parameter in PDA as it determines the extent to which the information provided by the proxy is weighted against prior information in the resulting reanalysis. This method provides a sound basis through which assimilated proxy records influence the reanalysis depending on the strength of their relationship to the dependent climate variables. For example, a record with a poor fit to calibration data will be characterized by larger residuals, hence larger observation error variance, and less weight in the reanalysis relative to a record that has a stronger correlation with climate variables. We note that modestly different results are obtained with different observational calibration datasets (see H16).

The calibration datasets used in this study are the NASA Goddard Institute for Space Studies (GISS) Surface Temperature Analysis (GISTEMP) (Hansen et al., 2010) version 4 for temperature, and the gridded precipitation dataset from the Global Precipitation Climatology Centre (GPCC) (Schneider et al., 2014) version 6 as the source of monthly information on moisture input over land surfaces.

2.4.1 Seasonality

Here we take advantage of the availability of expert information about the seasonal response to temperature for each proxy record included in the PAGES2k-2017 metadata. This information is not available in PAGES2k-2013, hence the use of PSMs calibrated on annual averages for all records in H16. Seasonality information is provided for each record as a numerical representation of a sequence of consecutive months (e.g. JJA as [6,7,8]). Seasonal PSMs are derived by using this sequence as the averaging period defining $\overline{X_1'}$ and $\overline{X_2'}$ in Eqs. 6 and 7.

Precise information on proxy seasonality is however not available for all records in the updated LMR proxy database. The proxies from Anderson et al. (2018) (section 2.2) have not been subjected to extensive community–wide screening and vetting as with the PAGES2k-2017 proxies. In particular, seasonality information for the large number of additional tree ring records from B14 has been encoded using a simple latitudinal dependence which does not attempt to represent possible record-by-record diversity (see Anderson et al., 2018). This lack of expert-informed seasonality motivates an objective alternative to the metadata seasonality information for calibrating tree ring width (TRW) forward models. We consider several potential seasonal periods, perform a regression over each possible season, and identify the linear relationship providing the best fit to proxy values, as defined by the maximum value of the adjusted R^2, a goodness-of-fit measure defined as (Goldberger, 1964, p. 217):

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(N-1)}{N-M-1}\right]. \tag{8}$$

Here, R^2 is the variance explained by the linear model, N is the sample size and M is the number of predictors in the model. The adjusted R^2 penalizes complexity (i.e. the number of predictors) of the model in such a way that values characterizing a more complex model will increase only if the additional predictors improve the fit more than would be expected by chance. Test periods considered include, in addition to the seasonal response in the proxy metadata (if available), the calendar year, boreal summer (JJA) and boreal winter (DJF), and extended Spring and Fall growing seasons (MAMJJA, JJASON for NH trees, SONDJF, DJFMAM for SH trees) to account for ecosystem–dependent variations in tree growth shifted toward the earlier or later parts of the warm season (see, e.g., Sano et al., 2009; D'Arrigo et al., 2005). With this test set of seasonal responses, the dominant sensitivity of some TRW chronologies to winter temperature (D'Arrigo et al., 2012) is included, as well as the winter and spring precipitation sensitivities characterizing some tree species (see, e.g., Stahle et al., 2009; Touchan et al., 2003). The latter point is germane to the calibration of seasonal TRW models using precipitation as a predictor (see next section).

2.4.2 Tree ring width sensitivity to temperature and moisture

Proxy number is strongly dominated by TRW records in the LMR proxy database, particularly with the addition of chronologies from B14. Furthermore, these records have not been screened on temperature, which opens the opportunity to measure moisture sensitivity through the regression framework. The addition of an explanatory variable increases the potential for overfitting, and our framework is designed to measure that using the 25% of proxies withheld from assimilation, for which we can measure reconstruction errors and compare results with proxies that were assimilated.

Two methods are considered, both adding a dependence to moisture input (as represented here with precipitation). The first maintains the univariate approach (Eq. 6) but considers linear PSMs calibrated against either temperature or precipitation. For each TRW record, distinct regressions with either variables are established and the model providing the best fit to proxy data is selected. Following a common practice in dendroclimatology, this approach determines whether the record is predominantly temperature or moisture limited (see, e.g., St. George, 2014). Similar univariate "temperature or moisture" models (abbreviated as "TorM" hereafter) are successfully used in Steiger et al. (2018). The second method consists of simultaneously factoring both temperature and moisture sensitivities through the bivariate relationship expressed in Eq. 7.

Seasonal univariate TorM and bivariate TRW models are considered, with distinct sets of models calibrated using proxy seasonality either from the proxy metadata or objectively-derived during calibration. This selection has important implications for the representation of the proxy seasonal response to moisture in particular. For the proxy metadata, seasonality for moisture is assumed to be identical to temperature as this is the only information available, whereas the objective approach allows for independent encoding of seasonal responses to temperature and moisture. For TorM models, the objective seasonality for univariate moisture models is independent of temperature as it is determined solely from the fit to precipitation data. For bivariate PSMs, all possible combinations of seasonal responses specified independently for temperature and moisture are considered and the combination providing the best fit is selected. With such flexibility, TRW models with objectively-derived seasonality are expected to provide a more realistic representation of the significant variability in seasonal responses to moisture characterizing TRW records (see, e.g., St. George et al., 2010). We note that this approach is similar to the methodology used to calibrate the VS-Lite model (Tolwinski-Ward et al., 2011) in that local temperature and precipitation data are used to determine site-specific growth seasons and seasonally-dependent temperature and moisture growth parameters.

An examination of PSM characteristics, summarized here, with more detail provided in appendix B, confirms that proxies are represented more accurately by seasonal models, particularly for tree-ring wood density and width records (see Table B1). Moreover, more-accurate fits to TRW data are obtained when proxy seasonal responses are determined objectively during model calibration. Finally, the addition of moisture input as a climate driver in TRW modeling proves most beneficial when implemented in bivariate models (see Table B2). These findings serve as the basis for defining a PDA configuration used for the reconstruction described in the next section.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

3 The updated reanalysis

We present a comparison between the updated reanalysis described by the method in the previous section with the LMR prototype described in H16[2]. Specifically, the updated reanalysis consists of all proxy records in the expanded database, using objectively-derived seasonal PSMs, with a bivariate formulation for all TRW proxies and univariate for all other proxy types. Covariance localization is applied with a 25000 km cut-off radius. In the next section we identify the sources of improvement that contribute to the increase in skill of the updated reconstruction. Results are evaluated against various twentieth century instrumental data and reanalyses using the Pearson correlation coefficient and the coefficient of efficiency (CE) (Nash and Sutcliffe, 1970).

Figure 2a shows a comparison of reconstructed global–mean temperature (GMT) between the prototype and updated reanalyses over the entire Common Era. Similar features are observed in the ensemble mean from both reanalyses, namely the cooling trend over most of the Common Era, followed by the industrial–era warming. Superimposed on these main trends, significant multidecadal to multicentennial variability characterize both reanalyses, including a cool period prior to the industrial warming, consistent with the Little Ice Age (LIA). Noticeable differences also exist between the reanalyses, such as the absence of the relatively warm period during 870–1000CE, representing the Medieval Climate anomaly (MCA), in the updated LMR. Weaker decadal-scale temperature variations largely characterize the updated reconstruction, particularly after 1000CE when more proxies are assimilated. Cooler conditions also prevail during the LIA in the prototype compared to the updated reanalysis, although verification against instrumental–era temperature analyses provides evidence that the prototype reanalysis is too cold during the early part of the instrumental period.

GMT verification results of the LMR ensemble mean against various instrumental temperature products are shown in Figs. 2c and d for the prototype and updated reanalyses respectively. Noticeably higher verification scores characterize the updated LMR, including a 10% increase in CE relative to the average of observations-based temperature analyses ("consensus"), and an increase in CE in the verification of the detrended GMT (over 1880-2000CE) from 0.32 in the prototype to 0.60 in the updated reanalysis (see Table 1). A narrower ensemble is also a characteristic of the new reanalysis, indicating a decrease in reconstructed GMT uncertainty. This represents an improvement over the prototype from the point of view of ensemble calibration, defined as the relationship between the mean squared error in the ensemble-mean and ensemble variance. The two should closely match for a well-calibrated ensemble. The GMT ensemble from the prototype is found to be overdispersive, meaning that the ensemble variance is much larger than the error in the ensemble mean, whereas the GMT ensemble in the updated LMR is found to be well-calibrated (see section 4.3 for more details).

Having access to ensembles can provide further insight into uncertainty in the reanalyses. It can be shown mathematically that the assimilation of observations invariably leads to a reduction in the variance characterizing the posterior ensemble compared to the prior. The ratio of ensemble variances of the posterior (reanalysis) to the prior is a measure of the information provided by the assimilated proxies. Figure 2b shows the temporal evolution of $1 - \mathrm{Var}[\mathbf{x}_a]/\mathrm{Var}[\mathbf{x}_b]$, so that a value of zero

[2]We use the experiment included in Figure 12 of H16, with PSMs calibrated using GISTEMP. Moreover, we use this configuration to generate a reconstruction of the Palmer Drought Severity Index (PDSI), which was not included in H16.

indicates no influence from proxies and one implies that all error has been removed. In the early part of the Common Era, when few proxy data are available, a variance decrease of only 10% occurs in the prototype compared to 20% for the updated reanalysis. The influence of proxies gradually increases after 450CE, at similar rates in both reanalyses. The reduction in variance is more pronounced in the updated LMR beginning at 1500CE, corresponding to the period with a significantly larger number of proxies (see Fig. 1). The largest reduction, 60% in the prototype compared to 80% to 90% in the updated reanalysis, is found during the 20th century when the most proxies are available. The uncertainty characterizing the updated reanalysis is therefore reduced overall compared to the prototype, which underscores the importance of the expanded proxy database in LMR.

Spatial verification is provided by comparing the LMR gridded 2m air temperature field against the Berkeley Earth (BE) instrumental–era temperature analysis (Rohde et al., 2013) (Fig. 3). BE is chosen as the verification reference as it has not been used to calibrate the PSMs, and provides the most complete spatial coverage compared to other instrumental products. The updated temperature reconstruction is largely improved compared to the prototype over large areas, including the tropical Pacific, northern Atlantic, North America, northern Europe, eastern and central Asia, Oceania, and over the Pacific sector of the Southern Ocean. The improvement is reflected in both correlation and CE scores, indicating improved timing and amplitude in reconstructed temperature variability. Exceptions are found over parts of the southern Atlantic and Indian oceans, although the decrease in skill is more modest compared to the magnitude of improvements elsewhere.

Hydroclimate verification is defined by a comparison of the reconstructed Palmer Drought Severity Index (PDSI) with the Dai (2011) product. We note here that this verification is entirely independent as TRW forward models were calibrated on precipitation and not on PDSI as in Steiger et al. (2018). PDSI is also improved in the updated reanalysis compared to the prototype (Fig. 4). Enhanced skill is particularly noticeable over North America, Europe and Asia, where most of the additional TRW records are located. An exception is the decreased skill found over a narrow band along the Siberian Taiga.

Finally we verify a climate variable away from the surface, the 500 hPa geopotential height field, against the corresponding field from NOAA's twentieth century reanalysis (20CR-v2, Compo et al., 2011) (Fig. 5). Once again we find the largest improvements over extratropical continental locations, and over the Arctic. We note similar improvements are found in the tropics and Northern Hemisphere mid-latitudes when verified against the ERA-20C reanalysis (Poli et al., 2016) (not shown). However, over the Northern and Southern Hemisphere high latitudes verification against ERA-20C is worse, which underscores significant differences between twentieth-century reanalyses in these data-sparse regions.

Table 1 summarizes the verification results discussed above through globally averaged verification scores. Improvements in the updated reanalyses are evident for all reconstructed variables, particularly with respect to the CE score, which is sensitive to bias and amplitude in interannual variability. These skill improvements suggest significant positive impact from the updated coral proxies and the addition of a large number of tree-ring proxies at higher latitudes. Furthermore, we anticipate that generalizing PSMs to accounting for seasonality and moisture sensitivity for TRW proxies also contribute to the improvements. In the following section we systematically evaluate improvements from these sources.

4 Sources of improvement

In this section, we turn our attention to the identification of the sources of reanalysis improvement. Results from multiple reconstruction experiments are presented, designed to quantify the impact of PSM formulation, the assimilation of various proxy data sets and the role of covariance localization.

4.1 Proxy system models

The different PSM configurations described in section 2.4 are used in a series of reconstruction experiments. In order to isolate PSM improvements, we first use PAGES2k-2017 proxies exclusively, as they represent the community standard in Common Era proxy information, and have well-defined seasonal metadata.

The impact of seasonal PSMs is first considered with three experiments performed using univariate temperature regression models for: (1) annual-mean calibration; (2) seasonality defined by expert metadata; and (3) objectively determined seasonality. Performance is again measured by correlation and CE scores with verification against the Berkeley Earth analysis. Relative to reconstructions with annual-mean PSMs (Figs. 6a and b), the reconstructions with seasonal PSMs (Figs. 6c–f) show improvements in both measures over nearly the entire globe (Figs. 6g–j). Results show a larger improvement for CE (Figs. 6h and j) compared to correlation (Figs. 6g and i), reflecting improvement in both the amplitude of temperature variability and bias. Noteworthy improvements are found in regions with large numbers of tree-ring proxies, such as the western United States, the region around and including Alaska, Northern Canada and western Arctic ocean, over Scandinavia and Norwegian Sea, central Asia and over the Southern Pacific west of the Antarctic Peninsula (see Fig. 6h). Comparing the differences of correlations and CE in Figs. 6i and j to those shown in Figs. 6g and h reveals that PSMs with objectively-derived seasonality contribute positively to skill for the aforementioned regions, especially where tree ring width records are most abundant (e.g. North America and Asia).

We turn now to the impact of moisture on seasonal TRW PSMs on the reconstructions. Since objectively defined seasonality performs best (i.e., Figs. 6e and f), reconstructions generated with univariate PSMs are used as the reference for measuring skill improvements for modeling TRW records as univariate in either temperature or moisture (abbreviated as “TorM”) (Figs. 7c and d) and for bivariate “temperature and moisture” PSMs (Figs. 7e and f). Improvement over univariate PSMs is apparent for the bivariate approach compared with the univariate “TorM” approach (cf. Fig. 7 panels g,h with i,j, respectively). In the bivariate approach regions such as western North America and central Asia, where most of the TRW records are found, improve the most in CE, but also over Australia, likely in response to the improved modeling of TRW records in New Zealand and Tasmania. Improvements are also noticeable through teleconnections with the central Atlantic and southern India Oceans, and over the eastern North Pacific Ocean. A decrease in skill is present over the mid-latitude Pacific ocean, but this is smaller in magnitude compared with skill enhancements elsewhere.

Verification of GMT for reconstructions using seasonal PSMs (Fig. 8) yields a similar interpretation to the spatial verification results. Compared to the consensus of instrumental–era products, we find that the 20th century trend in GMT is overestimated with the PAGES2k-2017 proxy data set if univariate PSMs are used. This is particularly the case with annual PSMs. Better agreement is obtained when seasonal bivariate PSMs are used to model TRW proxies. The representation of GMT interannual variability as measured by verification of the detrended GMT is also improved with seasonal PSMs, particularly for the CE metric. Similar to spatial verification results, PSMs with objectively-derived seasonality and bivariate TRW modeling have GMT reconstructions with consistently higher skill scores.

We recognize that the previous evaluation relies on comparisons with observation-based products covering the same time period as the data used to calibrate the statistical PSMs. To test the sensitivity of the results to the calibration period, we conduct additional independent instrumental–era calibration–validation experiments where PSMs are calibrated over a subset of the instrumental-era period and reconstructions are evaluated with data not used in calibration. Moreover, we perform an independent evaluation of reconstructions in proxy-space using proxies withheld from assimilation, for both the calibration and pre-calibration periods. These results, described in sections S3 and S4 in the supplementary material, confirm the main results and conclusions drawn here on the superiority of seasonal PSMs relative to those calibrated with annual averages, and the use of bivariate models for TRW proxies. Building upon this improved proxy modeling, we turn now to the role of increasing the number of proxy measurements on reconstruction skill.

4.2 Proxy data sets

Here we explore the role of the number of assimilated proxies on reanalysis verification. In order to measure this impact with the best configuration, the reconstruction experiments reported in this section are carried out using seasonal PSMs with objectively-derived seasonality for all records, with a bivariate formulation on temperature and precipitation for all TRW proxies, and univariate on temperature for all other proxies. The baseline reconstruction uses the PAGES2k-2017 proxies (as in the previous section), which we compare to experiments with the addition of the B14 TRW records, and then the further addition of the coral, ice and lake core records from Anderson et al. (2018).

Differences in correlation and CE associated with the addition of the B14 collection over the PAGES2k-2017 proxies show skill improvements over the continental United States and Mexico, Europe, and the southern edge of the Tibetan Plateau, where the additional records provide enhanced coverage (cf. Figs. 9g and h). Through the influence of significant spatial covariances with the added records, assimilation of the additional TRW records also leads to improved temperature skill over remote areas of the mid-latitude Pacific, northern Atlantic and Indian oceans.

The addition of records described in Anderson et al. (2018) has minimal additional impact overall, with the exception of modest increases in correlation and CE over Greenland (see see Figs. 9i and j). Results for GMT show that the 20th century trends from the prototype and the three experiments described here are within the uncertainty of the instrumental products (Fig. 10). However, the consensus trend (defined as the mean of trends from all instrumental analyses) is best reproduced with the most comprehensive proxy network. Skill metrics for the detrended GMT are higher for all of the updates to the proxy database compared to the prototype configuration, with a slight improvement for the reconstruction using the PAGES2k-2017 and B14 records; however, differences between results with the various subsets of the expanded proxy database are not statistically significant at the 95% confidence level.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

A comparison of the reconstructed PDSI between the prototype and experiments discussed in this section (Fig. 11) indicate that the increase in skill shown in Fig. 4 derives from the combined influence of the updated PAGES2k-2017 records and the B14 TRW records. With bivariate TRW forward models, enhanced skill is found predominantly over the western United States and Asia, as well as over eastern Europe (Figs.11g and h). The impact of adding the B14 TRW records is mostly found over the eastern part of the United states and over western Europe (Figs.11i and j). Finally, we note that the additional coral, ice and lake core records from Anderson et al. (2018) do not significantly affect the PDSI reconstruction skill (Figs.11e and f are nearly identical to Figs. 4c and d).

4.3 Covariance localization

A key question with ensemble data assimilation is whether spatial covariance localization should be applied, and if so, with what length scale (i.e. cut-off distance). Localization is applied to minimize the adverse impact of spurious covariances at large distances from a proxy location, which results from sample error in finite ensembles (Hamill et al., 2001). If localization is not applied, spurious covariances allow proxies to affect remote locations, which adversely affects the quality of the analysis. On the other hand, too-short localization length scales reduces the useful information that can be derived from the proxies. Therefore a balance is sought between minimizing sampling noise versus retaining useful proxy information.

We use the Gaspari-Cohn fifth-order polynomial with a specified cut-off radius (Gaspari and Cohn, 1999) for the localization function w_{loc} in (4). A series of reconstructions are performed with localization radii ranging from 5000 km to 25000 km. As with previous experiments, 51 Monte-Carlo realizations are carried out, each with 100 ensemble members assimilating 75% of proxy records in the expanded LMR database. In addition to the verification scores previously described, an additional discriminating factor on the quality of the reanalysis is "ensemble calibration" as defined by (Murphy, 1988),

$$ECR = \left[\frac{1}{N-1}\sum_{n=1}^{N}(v_n - \overline{x_n})^2\right]\left[\frac{1}{N-1}\sum_{n=1}^{N}(\sigma^2_{x,n} + \sigma^2_{v,n})\right]^{-1}, \tag{9}$$

where the numerator is the mean square error (MSE) of the analysis ensemble mean with respect to a verification data set v, and the denominator is the innovation variance: the sum of the analysis ensemble variance σ^2_x and the error variance σ^2_v characterizing the verification data. Here we apply Eq. 9 to the GMT ensembles from each Monte-Carlo realization, where the verification data corresponds to the consensus of instrumental–era data sets as before. The error variance σ^2_v is estimated from the deviations of the constituent sample data (the instrumental datasets) from their mean (i.e. the consensus), and the N "observations" are taken from the time series of annual analysis and verification data points over the 20th century. The ECR ratio expresses the degree to which the ensemble predicts the distribution of observations. A well-calibrated ensemble exhibits an approximate agreement between the ensemble variance and the ensemble–mean MSE, i.e. ECR $\approx$ 1.0, while an overdispersive ensemble has variance larger than the ensemble-mean MSE (ECR $<$ 1.0), and an underdispersive ensemble is diagnosed when its variance is smaller than the ensemble–mean MSE (ECR $>$1.0).

An ECR value is obtained for each Monte-Carlo reconstruction, and the average is taken over the 51 values. Results indicate that GMT ensembles tend to be underdispersive without covariance localization (Table 2). An excessive reduction in posterior ensemble spread, compared to the mean error, can arise from two factors: too-small observation error variance for some records,

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

leading to overestimated weight of these records in the reanalysis, combined with sample error in the ensemble-estimated covariances, which artificially reduces ensemble variance over the entire state vector. In contrast, ensembles are overdispersive when "too-small" localization radii are used, indicating the information on global-mean temperature provided by some proxy records is not properly incorporated in the reanalysis. The ECR closest to one (i.e. well-calibrated ensemble) is obtained when covariance localization is applied with a cut-off radius of about 25000 km. This result, along with those from the common verification scores (see Fig. S4), suggest that a skillful and reliable reconstruction is obtained with this covariance localization configuration.

We note that the optimal localization radius depends on a number of factors, such as ensemble size, the observation network and observation error characteristics. For example, as seen in Table 2, an overdispersive GMT ensemble is obtained when proxies from PAGES2k-2013 are assimilated in the absence of covariance localization, due to the much smaller number of proxies. This is in contrast with the underdispersive ensemble resulting from the assimilation of the significantly larger number of proxies from the updated proxy database used here.

5 Concluding summary

A paleoclimate reanalysis of the Common Era has been developed using an updated data assimilation framework. Results show significant improvement over the prototype Last Millennium Reanalysis presented in Hakim et al. (2016). The development of a vastly expanded proxy database and implementation of proxy system models (PSMs) with improved realism are shown to be key contributors to the enhanced reanalysis. Upgrades to the proxy database consist of a change from the community-standard of PAGES 2k Consortium (2013) to the more recent PAGES 2k Consortium (2017) data set, complemented by the records described in Anderson et al. (2018), bringing a fivefold increase in the number of proxy records available for assimilation. Moreover, new methods to map state variables to observations extend the prototype's linear univariate models calibrated on annual-mean temperature in two key aspects: accounting for seasonal dependencies of individual proxy records, and the modeling of tree-ring-width proxies using temperature and moisture as predictors. The encoding of proxy seasonality information within PSMs has also been refined by objectively determining the characteristic seasonal response of individual records, and by decoupling the seasonality for temperature and precipitation sensitivity for tree-ring-width.

Climate field reconstructions from a series of assimilation experiments carried out with various proxy and PSM configurations have been compared to available instrumental–era observation-based analyses, revealing notable improvements not only in the reconstructed global mean temperature in general, but also in reconstructed spatial fields. More skillful tropical Pacific temperatures are obtained primarily due to the updated set of coral records in the PAGES 2k Consortium (2017) collection. Improved temperature reconstructions over continental extratropical regions are the result of the newly implemented seasonal PSMs, combined with the assimilation of the large number of Breitenmoser et al. (2014) tree-ring-width chronologies, forward-modeled using a bivariate temperature-moisture formulation. Improvements are reflected not only in temperature reconstructions, but also in dynamical variables (500 hPa geopotential heights) and in hydroclimate variables such as the PDSI. The introduction within LMR's proxy database of the large collection of tree-ring-width records not screened for temperature

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

Climate of the Past Discussions
Open Access

sensitivity appears to be a significant factor in enabling this capability. Lastly, covariance localization, applied with a relatively large cut-off length scale, has been shown to have a positive impact, particularly in maintaining an appropriate relationship between mean error and variance in the reanalysis ensemble.

Results presented here, based upon regression PSMs, may serve as a reference for future efforts designed to assess the value of more comprehensive process-based PSMs in paleoclimate data assimilation research. Finally, we note that the version of the PDA system described herein corresponds to the configuration used in the production of the first publicly released NOAA Last Millennium Reanalysis, available at https://atmos.washington.edu/~hakim/LMR/.

Code and data availability. The code used in the production of the reanalysis is publicly available at https://github.com/modons/LMR, and data are available from https://atmos.washington.edu/~hakim/LMR/.

Appendix A: DA with an appended state

For reasons of computational efficiency and flexibility we perform data assimilation with an "appended state", where the $\mathbf{y}_e$ proxy estimates from each record are appended to the state vector $\mathbf{x}_b$:

$$\mathbf{x}_b = \begin{bmatrix} x_1 \\ \vdots \\ x_N \\ y_e^1 \\ \vdots \\ y_e^P \end{bmatrix}, \tag{A1}$$

where the $x_1 \ldots x_N$ elements contain the ensemble grid point data from model variables included in the state (e.g. temperature, precipitation etc.), with N the sum of the number of variables times the number of grid points, and the $y_e^1 \ldots y_e^P$ are the ensemble proxy estimates for each of the P proxy records considered. Each of the $x_1 \ldots x_N$ and $y_e^1 \ldots y_e^P$ elements are of dimensions $1 \times N_{ens}$, where N_{ens} is the specified size of the ensemble. Hence, $\mathbf{x}_b$ is a matrix of dimension $(N+P) \times N_{ens}$. With such an appended state, the y_e elements in Eq. A1 are updated through Eqs. 4 as any other state variables, eliminating the need to re-evaluate $\mathbf{y}_e$ with Eq. 2 once the state has been updated. This simplification is particularly attractive in the context of LMR updates discussed herein as it enables a straightforward implementation of seasonal PSMs (i.e. forward models more accurately representing the seasonal responses of individual proxy records) as discussed in section 2.4. In our implementation with an appended state and serial processing of observations, along with the reconstruction of annually-averaged states, the data assimilation procedure follows this general algorithm.

1. The proxy estimates ($\mathbf{y}_e$) are pre-calculated using Eq. 2 with either annually- or seasonally-averaged model data as input (i.e. the $\mathbf{x}_b$ in Eq. 2).

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

2. A sample ensemble of annually-averaged model states is randomly drawn from a pre-existing simulation to form the main part of the prior state vector (i.e. the $x_1 \dots x_N$ elements in Eq. A1).

3. The pre-calculated $y_e^1 \dots y_e^P$ proxy estimates are added on to form the appended state as shown in Eq. A1. This appended state becomes the $\mathbf{x}_b$ in Eq. 1, which is decomposed in an ensemble-mean ($\overline{\mathbf{x}}_b$) and perturbations about the mean ($\mathbf{x}'_b$) as shown in Eqs. 4.

4. Proxies forming the $\mathbf{y}$ vector are then serially processed, with the updated state, including the proxy estimates, obtained from Eqs. 4. The complete reanalysis is completed once all proxies have been assimilated.

We note here that with a configuration involving seasonal PSMs without the use of an appended state, the vector $\mathbf{x}_b$ has to include states with sufficient temporal resolution to allow the calculation of the updated seasonal $y_e^1 \dots y_e^P$ proxy estimates. In this scenario, an additional step to the ones listed above is required, involving Eq. 2 using the appropriate seasonally-averaged updated states as input. With proxies characterized by a wide range of seasonal responses, this requirement would impose an $\mathbf{x}_b$ composed of monthly data which would greatly increase the computational cost of the reanalysis. Reanalysis results would also likely be adversely affected by the larger noise level characterizing data at shorter (i.e. monthly) timescales through its impact on ensemble estimates of prior covariances (see, e.g. Tardif et al., 2016).

Appendix B: Proxy system model characteristics

Features introduced in the updated LMR proxy modeling capabilities include a representation of the seasonal response to climate drivers characterizing individual proxy records (i.e. proxy seasonality), as well as proxy system models (PSMs) that include moisture input and temperature as driving variables for modeling tree-ring-width (TRW) records.

The first approach is to use univariate PSMs calibrated against temperature data, with proxy seasonality either defined from the available proxy metadata or derived objectively using the method described in section 2.4.1. PSM performance is compared using the Bayesian Information Criterion (BIC), defined as (Schwarz, 1978),

$$BIC = -2\ln(\hat{L}) + k\ln(n) \tag{B1}$$

where $\hat{L}$ is the maximized value of the likelihood function of the model, n is the sample size and k is the number of estimated parameters in the model. We note that the second term in Eq. B1 represents a penalty for models with a larger number of explanatory variables, i.e. a more complex model. This feature is particularly useful when comparing univariate and bivariate models. Here we use the difference in BIC values between two models $\Delta BIC = (BIC_M - BIC_{ref})$, to determine the relative accuracy of model M over a reference. The model with the lowest BIC is preferred (i.e. a better fit to the data), hence a negative ΔBIC indicates the superiority of the test model over its reference. Here, the seasonal PSMs are tested against the univariate PSMs calibrated with annually-averaged temperatures as the reference. Significant evidence of the superiority of the test model over its reference is obtained when ΔBIC < -2.0.

Table B1 presents a summary of ΔBIC results for records in each proxy category considered in LMR. The advantage of seasonal PSMs is particularly significant for tree-ring wood-density chronologies, a proxy known for its strong seasonal response (Briffa et al., 2004). Seasonal PSMs also provide improved fits to tree ring width data, although to a lesser extent compared to density records. As indicated by the larger negative ΔBIC values, models based on objectively-derived seasonal responses lead to more accurate descriptions of proxy data compared to those calibrated using metadata seasonality, even for tree ring chronologies within the community-curated PAGES2k-2017 data set. These results suggest that the objectively-derived seasonality information is noticeably different than in the metadata, particularly for tree ring records in the Breitenmoser et al. (2014) (i.e. B14) data set, but also for those in PAGES 2k Consortium (2017) (i.e. PAGES2k-2017). More details on this aspect are provided in the supplementary material. The use of objectively-defined seasonality improves upon the simple latitude-dependent relationship described in Anderson et al. (2018), more consistent with records from the PAGES2k-2017 data set. Apart from lake sediment records, which are also more accurately modeled with seasonal PSMs, Table B1 shows that PSMs for other proxy types are not as sensitive to seasonality. In fact, the majority of the (tropical) coral records included in the current database have metadata seasonality defined as annual already, as do the high-latitude ice core records. Note that some of these records originate from the collection described by Anderson et al. (2018), where seasonality metadata information is generally not available. As a result, these records are assumed to be annual.

In addition to seasonal models, other improvements involve the development of PSMs which add precipitation as an input variable for the modeling of TRW proxies as outlined in section 2.4.2. One approach consists of selecting the univariate models, either calibrated on temperature or moisture input, which best describe the proxy data. This "temperature or moisture" selection (abbreviated as "TorM") is performed on individual TRW records, and the resulting proportion of TRW proxies identified as temperature-sensitive is 56.4% versus 43.6% for moisture when metadata seasonality information is considered. This is compared to 36.8% temperature-sensitive versus 63.2% moisture-sensitive trees when seasonal responses are determined objectively. The latter option, leading to a larger proportion of moisture-sensitive records, is in better agreement with a comparable characterization performed by Steiger et al. (2018) on a similar set of TRW records.

A second approach consists of bivariate PSM formulation, where TRW depends on both temperature and precipitation (see Eq. 7. The ΔBIC results characterizing the univariate "TorM" and bivariate PSMs against their univariate temperature-only counterparts (as the reference) are summarized in Table B2. The negative mean ΔBIC values confirm the advantage of including moisture in TRW linear models. The evidence is more pronounced for the B14 records, perhaps not surprisingly given the larger proportion of moisture-sensitive records included in this data set. Nonetheless, the prevalent reduction in BIC for models of PAGES2k-2017 trees suggests a non-negligible response to moisture despite the screening of records for temperature. The mean positive ΔBIC characterizing the bivariate models calibrated using metadata seasonality confirm that the assumption of identical seasonal responses for temperature and moisture is problematic for modeling tree ring growth, at least with these more complex models. On the other hand, allowing distinct representations of temperature and moisture seasonal responses in bivariate PSMs, as enabled by the goodness-of-fit objective determination of these responses, leads to significantly more accurate TRW modeling compared to univariate temperature PSMs.

Author contributions. R.T. developed the code related to all the updates discussed in the paper (generation of expanded proxy database, calibration and application of the new proxy forward models), configured and performed all the reanalysis experiments, analyzed the experimental results and wrote the manuscript with G.J.H. G.J.H. proposed the general design of the updated tree-ring proxy models, contributed the Kalman update and original versions of the verification codes. W.A.P. developed several key parts of the reanalysis code. K.A.H. collected and formatted the proxy records taken from sources other than the PAGES 2k collection, re-processed the tree ring data taken from Breitenmoser et al. M.P.E. contributed to the identification of proxy records to be included/excluded from the updated proxy database and suggested key code modifications. J.E.G. provided the PAGES 2k proxy data and advice on proxy system modeling. D.M.A. supervised the collection of proxy records and helped define the content of the proxy database. E.J.S. contributed advise on proxy system modeling, statistical measures, and helped define the content of the proxy database. D.N. contributed advice on proxy system modeling. All authors contributed to the final editing of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank Nathan Steiger of the Lamont-Doherty Earth Observatory, Columbia University, for providing the PDSI data from the CCSM4 Last Millennium simulation, and Jörg Franke of the University of Bern, Switzerland, for providing the tree-ring data from the Breitenmoser et al. (2014) study. This research was supported by grants from the National Oceanic and Atmospheric Administration (grant NA14OAR4310176) and the National Science Foundation (grant AGS-1304263 to the University of Washington) and . We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. CMIP5 data used in this paper may be obtained from the Earth System Grid Federation at http://esgf.llnl.gov/. Some calibration and verification data sets were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at https://www.esrl.noaa.gov/psd/.

References

Acevedo, W., Fallah, B., Reich, S., and Cubasch, U.: Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model, Clim. Past, 13, 545–557, https://doi.org/10.5194/cp-13-545-2017, 2017.

Anderson, D. M., Tardif, R., Horlick, K., Erb, M. P., Hakim, G. J., Emile-Geay, J., Noone, D., Perkins, W. A., and Steig, E. J.: Additions to the Last Millennium Reanalysis multi-proxy database (LMRdb v1.0.0), Data Science Journal, submitted, 2018.

Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, Clim. Past, 8, 963–976, https://doi.org/10.5194/cp-8-963-2012, 2012.

Breitenmoser, P., Brönnimann, S., and Frank, D.: Forward modelling of tree-ring width and comparison with a global network of tree-ring chronologies, Clim. Past, 10, 437–449, https://doi.org/10.5194/cp-10-437-2014, 2014.

Briffa, K. R., Osborn, T. J., and Schweingruber, F. H.: Large-scale temperature inferences from tree rings: a review, Glob. Planet. Change, 40, 11–26, https://doi.org/10.1016/S0921-8181(03)00095-X, 2004.

Cobb, K. M., Charles, C. D., Cheng, H., and Edwards, R. L.: El Niño/Southern Oscillation and tropical Pacific climate during the last millennium, Nature, 424, 271–276, https://doi.org/10.1038/nature01779, 2003.

Compo, G., Whitaker, J., Sardeshmukh, P., Matsui, N., Allan, R., Yin, A., Gleason, B., Vose, R., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R., Grant, A., Groisman, P., Jones, P., Kruk, M., Kruger, A., Marshall, G., Maugeri, M., Mok, H., Nordli, Ø., Ross, T., Trigo, R., Wang, X., Woodruff, S., and Worley, S.: The twentieth century reanalysis project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

Dai, A.: Characteristics and trends in various forms of the palmer drought severity index during 1900–2008, Journal of Geophysical Research: Atmospheres, 116, 1–26, https://doi.org/10.1029/2010JD015541, 2011.

D'Arrigo, R., Mashig, E., Frank, D., Wilson, R., and Jacoby, G.: Temperature variability over the past millennium inferred from Northwestern Alaska tree rings, Clim. Dyn., 44, 227–236, https://doi.org/10.1007/s00382-004-0502-1, 2005.

D'Arrigo, R., Anchukaitis, K. J., Buckley, B., Cook, E., and Wilson, R.: Regional climatic and North Atlantic Oscillation signatures in West Virginia red cedar over the past millennium, Global and Planetary Change, 84-85, 8–13, https://doi.org/10.1016/j.gloplacha.2011.07.003, 2012.

Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., and Thompson, D. M.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, Journal of Advances in Modeling Earth Systems, 7, 1220–1247, https://doi.org/10.1002/2015MS000447, 2015.

Dee, S., Steiger, N. J., Emile-Geay, J., and Hakim, Gregory J.: The utility of proxy system modeling in estimating climate states over the Common Era, Journal of Advances in Modeling Earth Systems, 8, 1164–1179, https://doi.org/10.1002/2016MS000677, 2016.

Emile-Geay, J., Cobb, K. M., Mann, M. E., and Wittenberg, A. T.: Estimating central equatorial Pacific SST variability over the past millennium. Part I: Methodology and validation, Journal of Climate, 26, 2302–2328, 2013.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Science Reviews, 76, 16–28, 2013.

Franke, J., Brönnimann, S., Bhend, J., and Brugnara, Y.: A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations, Sci. Data, 4, https://doi.org/10.1038/sdata.2017.76, 2017.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quart. J. Roy. Meteor. Soc., 125, 723–757, https://doi.org/10.1002/qj.49712555417, 1999.

Goldberger, A. S.: Econometric Theory, Wiley, New York, NY, 1964.

Goosse, H.: An additional step toward comprehensive paleoclimate reanalyses, Journal of Advances in Modeling Earth Systems, 8, 1501–1503, https://doi.org/10.1002/2016MS000739, 2016.

Goosse, H., Renssen, H., Timmermann, A., Bradley, R. S., and Mann, M. E.: Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium, Climate Dynamics, 27, 165–184, https://doi.org/10.1007/s00382-006-0128-6, 2006.

Goosse, H., E., C., de Montety, A., Mann, M. E., Renssen, H., and Timmermann, A.: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation, J. Geophys. Res.-Atmos., 115, https://doi.org/10.1029/2009jd012737, 2010.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N. J., and Perkins, W. A.: The Last Millennium Climate Reanalysis Project: Framework and First Results, Journal of Geophysical Research: Atmospheres, 121, 6745–6764, https://doi.org/10.1002/2016JD024751, 2016.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Monthly Weather Review, 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2, 2001.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Rev. Geophys., 48, https://doi.org/10.1029/2010RG000345, 2010.

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., and Teng, H.: Last millennium climate and its variability in CCSM4, Journal of Climate, 26, 1085–1111, https://doi.org/10.1175/JCLI-D-11-00326.1, 2013.

Liu, H., Liu, Z., and Lu, F.: A systematic comparison of particle filter and EnKF in assimilating time–averaged observations, Journal of Geophysical Research: Atmospheres, 122, https://doi.org/10.1002/2017JD026798, 2017.

Matsikaris, A., Widmann, M., and Jungclaus, J. H.: On-line and off-line data assimilation in palaeoclimatology: a case study, Clim. Past, 11, 81–93, https://doi.org/10.5194/cp-11-81-2015, 2015.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD017187, 2012.

Murphy, J. M.: The impact of ensemble forecasts on predictability, Quart. J. Roy. Meteor. Soc., 114, 463–493, 1988.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, Journal of Hydrology, 10, 282–290, 1970.

Okazaki, A. and Yoshimura, K.: Development and evaluation of a system of proxy data assimilation for paleoclimate reconstruction, Clim. Past, 13, 379–393, https://doi.org/10.5194/cp-13-379-2017, 2017.

PAGES 2k Consortium: Continental-scale temperature variability during the past two millennia, Nature Geoscience, 6, 339–346, https://doi.org/10.1038/ngeo1834, 2013.

PAGES 2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, Sci. Data, 4, https://doi.org/10.1038/sdata.2017.88, 2017.

Perkins, W. A. and Hakim, G. J.: Reconstructing paleoclimate fields using online data assimilation with a linear inverse model, Clim. Past, 13, 421–436, https://doi.org/10.5194/cp-13-421-2017, 2017.

Poli, P., Hersbach, H., and Dee, D. P.: ERA-20C: An atmospheric reanalysis of the twentieth century, J. Climate, 29, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1, 2016.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickman, C., and Mosher, S.: Berkeley Earth temperature averaging process, Geoinformatics and Geostatistics: An Overview, 1:2, 1–13, https://doi.org/10.4172/2327-4581.1000103, 2013.

Sano, M., Furuta, F., and Sweda, T.: Tree-ring-width chronology of *Larix gmelinii* as an indicator of changes in early summer temperature in east-central Kamchatka, J. Forest Research, 14, 147–154, https://doi.org/10.1007/s10310-009-0123-y, 2009.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theoretical and Applied Climatology, 115, 15–40, https://doi.org/10.1007/s00704-013-0860-x, 2014.

Schwarz, G. E.: Estimating the dimension of a model, Annals of Statistics, 6, 461–464, https://doi.org/10.1214/aos/1176344136, 1978.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880-2006), Journal of Climate, 21, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1, 2008.

St. George, S.: An overview of tree-ring width records across the Northern Hemisphere, Quaternary Science Reviews, 95, 132–150, https://doi.org/10.1016/j.quascirev.2014.04.029, 2014.

St. George, S., Meko, D. M., and Cook, E. R.: The seasonality of precipitation signals embedded within the North American Drought Atlas, The Holocene, 20, 983–988, https://doi.org/10.1177/0959683610365937, 2010.

Stahle, D. W., Cleaveland, M. K., Grissino-Mayer, H. D., Griffin, R. D., Fye, F. K., Therell, M. D., Burnette, D. J., Meko, D. M., and Villanueva Diaz, J.: Cool- and warm-season precipitation reconstructions over western New Mexico, J. Climate, 22, 3729–3750, https://doi.org/10.1175/2008JCLI2752.1, 2009.

Steiger, N. J. and Smerdon, J. E.: A pseudoproxy assessment of data assimilation for reconstructing the atmosphere–ocean dynamics of hydroclimate extremes, Clim. Past, 13, 1435–1449, https://doi.org/10.5194/cp-13-1435-2017, 2017.

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of time-averaged pseudoproxies for climate reconstruction, Journal of Climate, 27, 426–441, https://doi.org/10.1175/JCLI-D-12-00693.1, 2014.

Steiger, N. J., Steig, E. J., Dee, S. G., Roe, G. H., and Hakim, G. J.: Climate reconstruction using data assimilation of water isotope ratios from ice cores, Journal of Geophysical Research: Atmospheres, 122, 1545–1568, https://doi.org/10.1002/2016JD026011, 2017.

Steiger, N. J., Smerdon, J. E., Cook, E. R., and Cook, B. I.: A reconstruction of global hydroclimate and dynamical variables over the Common Era, Sci. Data, 5, https://doi.org/10.1086/sdata.2018.86, 2018.

Tardif, R., Hakim, G. J., and Snyder, C.: Coupled atmosphere–ocean data assimilation experiments with a low-order model and CMIP5 model data, Climate Dynamics, 45, 1415–1427, 2016.

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

Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., and Anchukaitis, K. J.: An efficient forward model of the climate controls on interannnual variation in tree-ring width, Climate Dynamics, 36, 2419–2439, https://doi.org/10.1007/s00382-011-1062-9, 2011.

Touchan, R., Garfin, G., Meko, D., Funkhouser, G., Erkan, N., Hughes, M., and Wallin, B.: Preliminary reconstructions of spring precipitation in southwestern Turkey from tree-ring width, Int. J. Climatol., 23, 157–171, https://doi.org/10.1002/joc.850, 2003.

Wang, J., Emile-Geay, J., McKay, N. P., and Rajaratnam, B.: Fragility of reconstructed temperature patterns over the common era: Implications for model evaluation, Geophysical Research Letters, 42, 7162–7170, https://doi.org/10.1002/2015GL065265, 2015.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Monthly Weather Review, 130, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2, 2002.

Widmann, M., Goosse, H., Schrier, G., Schnur, R., and Barkmeijer, J.: Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium, Clim. Past, 6, 627–644, https://doi.org/10.5194/cp-6-627-2010, 2010.

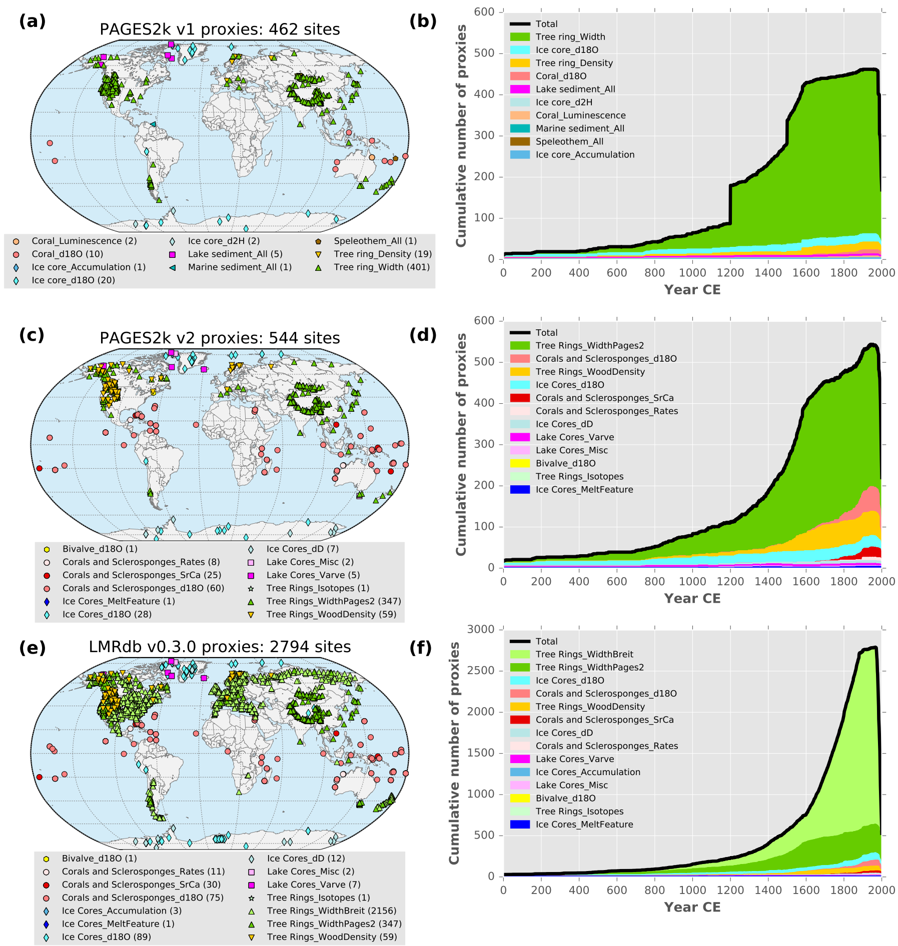

Figure 1. Locations (left column) and temporal (right column) distributions of proxy records available for assimilation (proxies for which linear PSMs calibrated with GISTEMP version 4 are available), (a) and (b) used in the prototype version, (c) and (d) LMR proxy database updated to PAGES 2k Consortium (2017) proxies, and finally (e) and (f) with the further addition of proxies from Anderson et al. (2018).

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

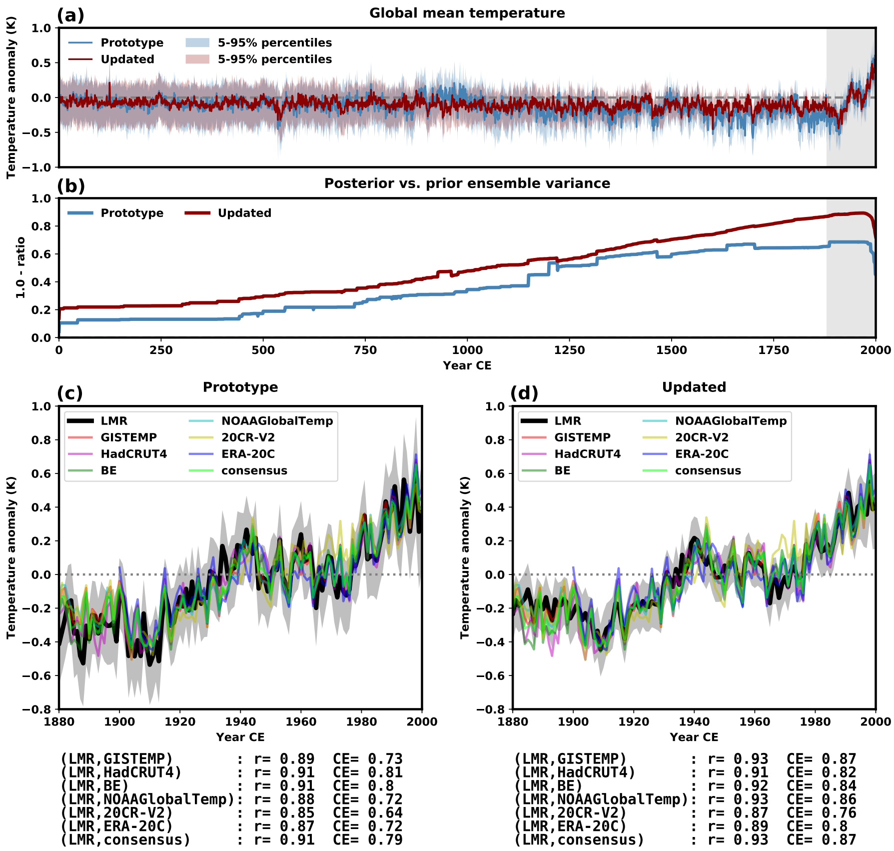

Figure 2. Comparison of the LMR global-mean 2 m air temperature (GMT) (a) grand ensemble mean (solid lines) and 5–95% percentile range (shading) from the prototype (blue) and updated (red) reanalyses, and (b) one minus the mean (across Monte-Carlo realizations) ratio of the posterior and prior GMT ensemble variance, over the Common Era. Comparison of LMR (c) prototype and (d) updated reanalyses, against instrumental–era analyses (GISTEMP: NASA GISS surface temperature (Hansen et al., 2010); HadCRUT4: Hadley Center/Climate Research Unit at the University of East Anglia temperature data set version 4 (Morice et al., 2012); BE: Berkeley Earth surface temperature (Rohde et al., 2013); NOAAGlobalTemp:NOAA merged land-ocean surface temperature version 3.5.4 (Smith et al., 2008); 20CR-V2: NOAA twentieth century reanalysis version 2 (Compo et al., 2011); ERA-20C: ECMWF reanalysis of the twentieth century (Poli et al., 2016); Consensus: average of all but LMR). The gray bands show the LMR 5–95% percentile range. Verification correlation (r) and coefficient of efficiency (CE) values are shown below the figures. The light gray shading in (a) and (b) indicate the verification period shown in (c) and (d).

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

Climate of the Past Discussions

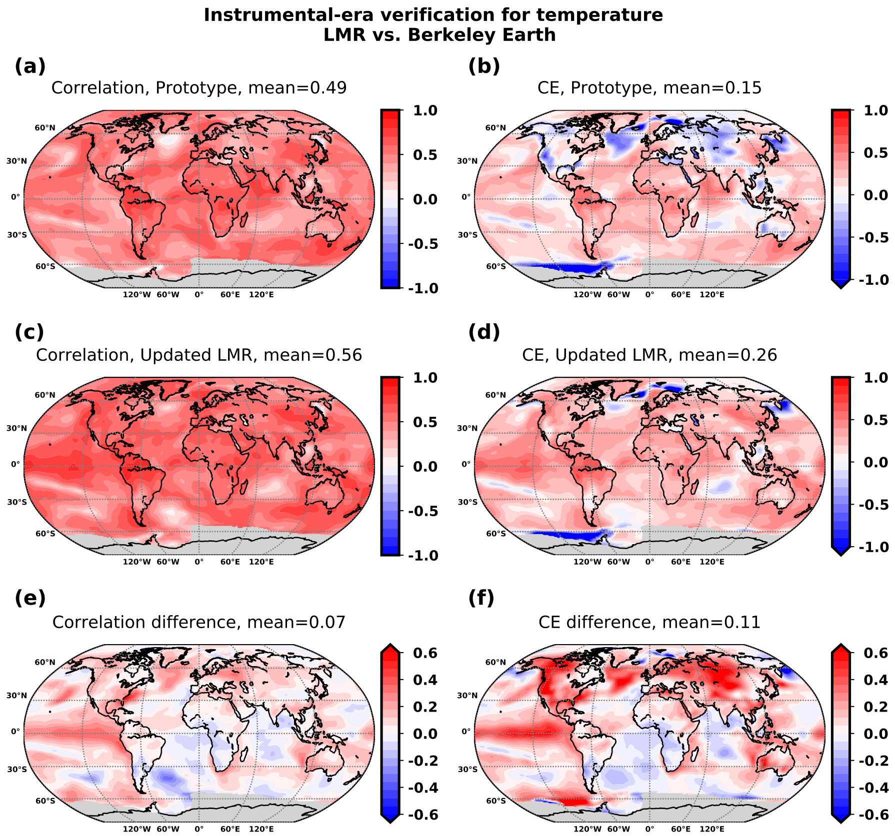

Figure 3. Verification of LMR 2m air temperature against the Berkeley Earth instrumental–era analysis over the 1880–2000 period. Shown are time series correlation (left column) and coefficient of efficiency (CE, right column), for (a) and (b) the prototype and, (c) and (d) the updated reanalysis. Differences in correlations and CE between the two experiments are shown in (e) and (f) respectively. Gray shading indicate regions with insufficient valid data for meaningful verification statistics.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

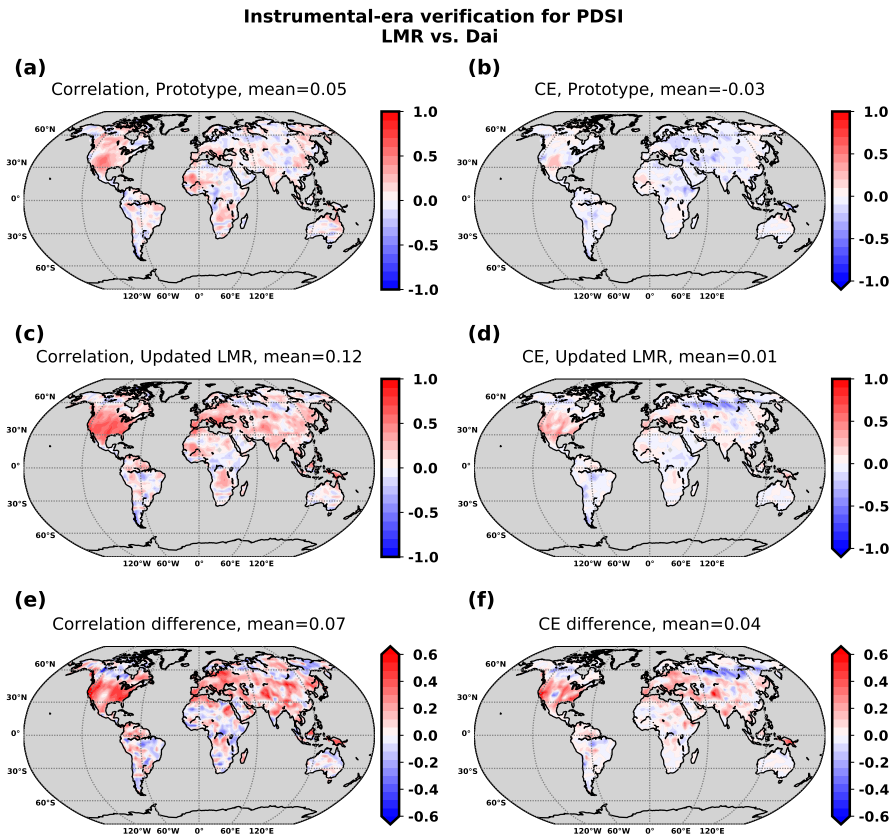

Figure 4. As in Fig. 3, except for verification of LMR Palmer Drought Severity Index (PDSI) against the Dai (2011) instrumental–era PDSI analysis.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

Climate of the Past Discussions
Open Access

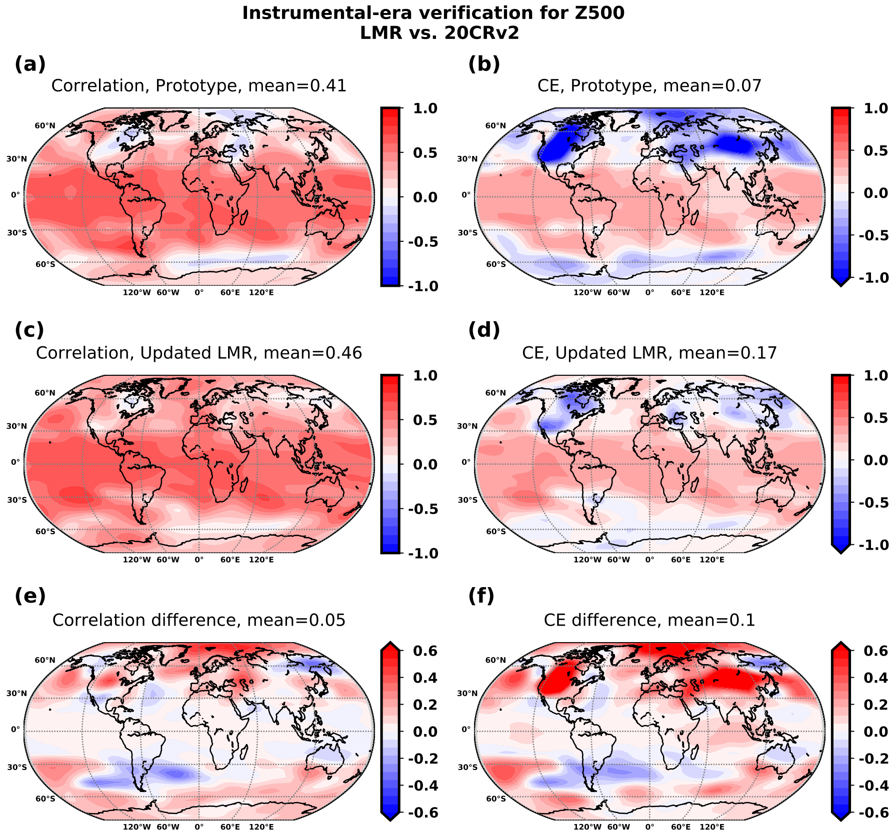

Figure 5. As in Fig. 3 except for the verification of LMR 500 hPa geopotential height anomalies against the 20CR-v2 reanalysis.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

Climate of the Past Discussions
Open Access

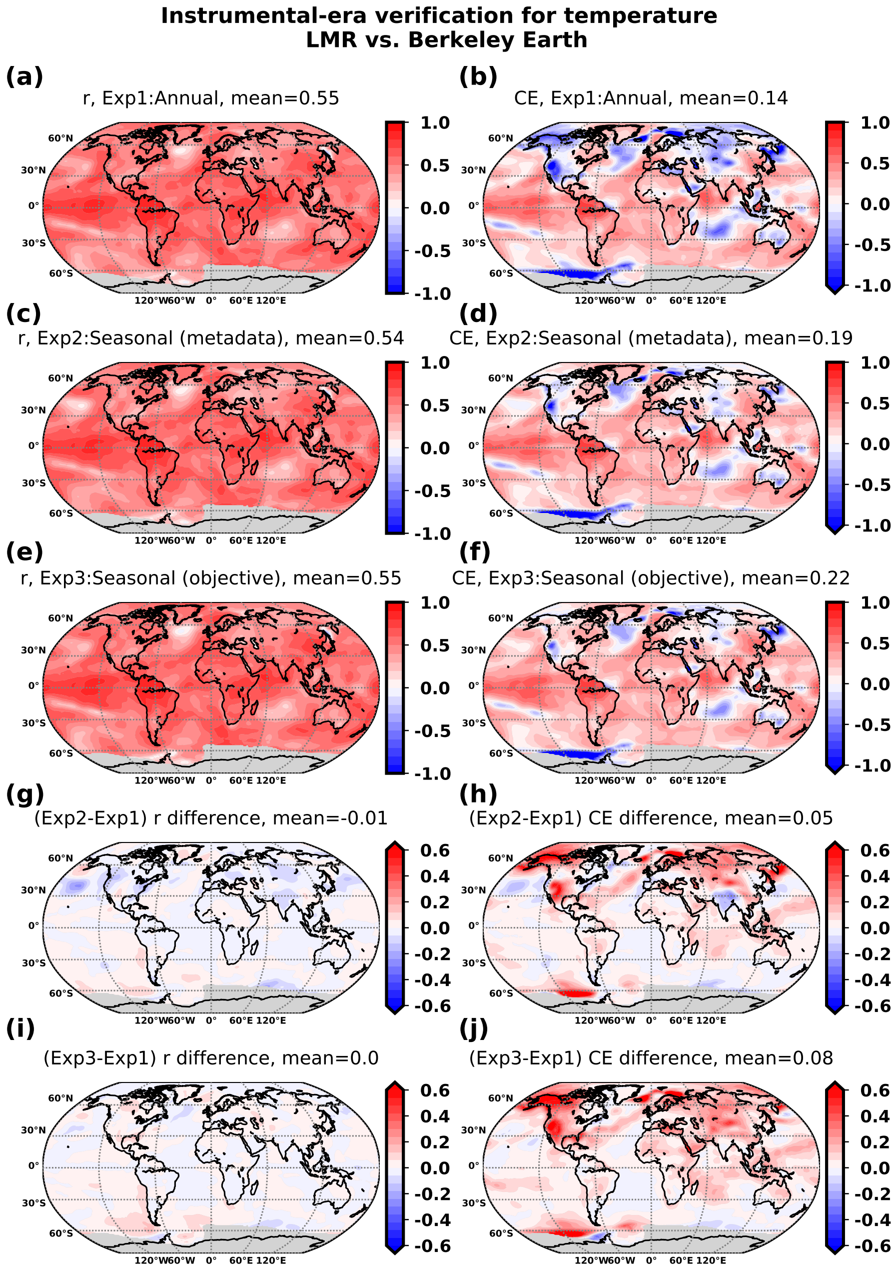

Figure 6. Verification of LMR temperature anomalies against the Berkeley Earth instrumental–era analysis, for experiments using PAGES2k-2017 proxies and univariate PSMs, with contrasting seasonalities. Shown are time series correlation (r) and coefficient of efficiency (CE), for (a) and (b) experiment 1: annual, (c) and (d) experiment 2: seasonality from the proxy metadata, and (e) and (f) experiment 3: objectively-derived seasonality. Differences in skill metrics are also shown, (g) and (h) between experiments 2 and 1, (i) and (j) between experiments 3 and 1.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

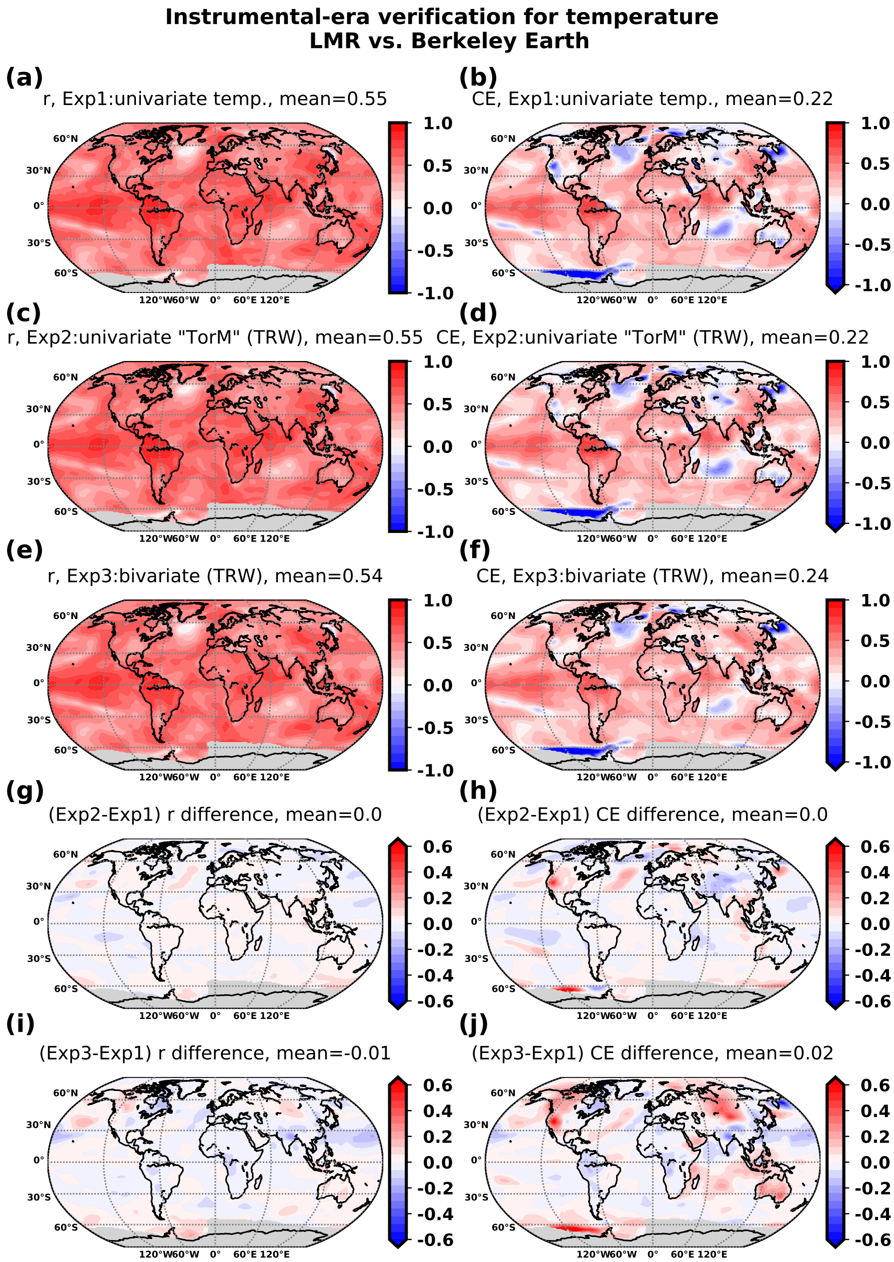

Figure 7. As in Figure 6, but comparing experiments performed using PAGES2k-2017 proxies with different PSM configurations for tree ring width proxies. (a) and (b) experiment 1: univariate on temperature for all proxies, (c) and (d) experiment 2: univariate with respect to temperature or moisture for TRWs, and (e) and (f) experiment 3: bivariate on temperature and moisture for tree ring widths. Differences in skill metrics are shown, (g) and (h) between experiments 2 and 1, and (i) and (j) between experiments 3 and 1. All reconstructions are based on objectively-derived seasonal PSMs.

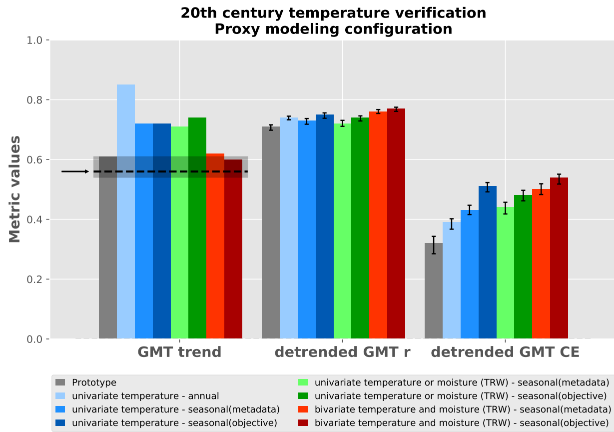

Figure 8. Summary skill metrics from verification of LMR temperature reconstruction experiments, all using PAGES2k-2017 proxies only, against the consensus of instrumental–era products. Experiments are a combination of those shown in Figs. 6 and 7. Metrics shown are the 20th century trend of global mean temperature (GMT), correlation (r) and coefficient of efficiency (CE) for the detrended GMT. The GMT trend from consensus of instrumental–era products is shown by the arrow and dashed black line, along with the range defined by the individual instrumental–era products shown by the gray-shaded area. Error bars are the 5-95% bootstrap confidence intervals on the corresponding skill metric.

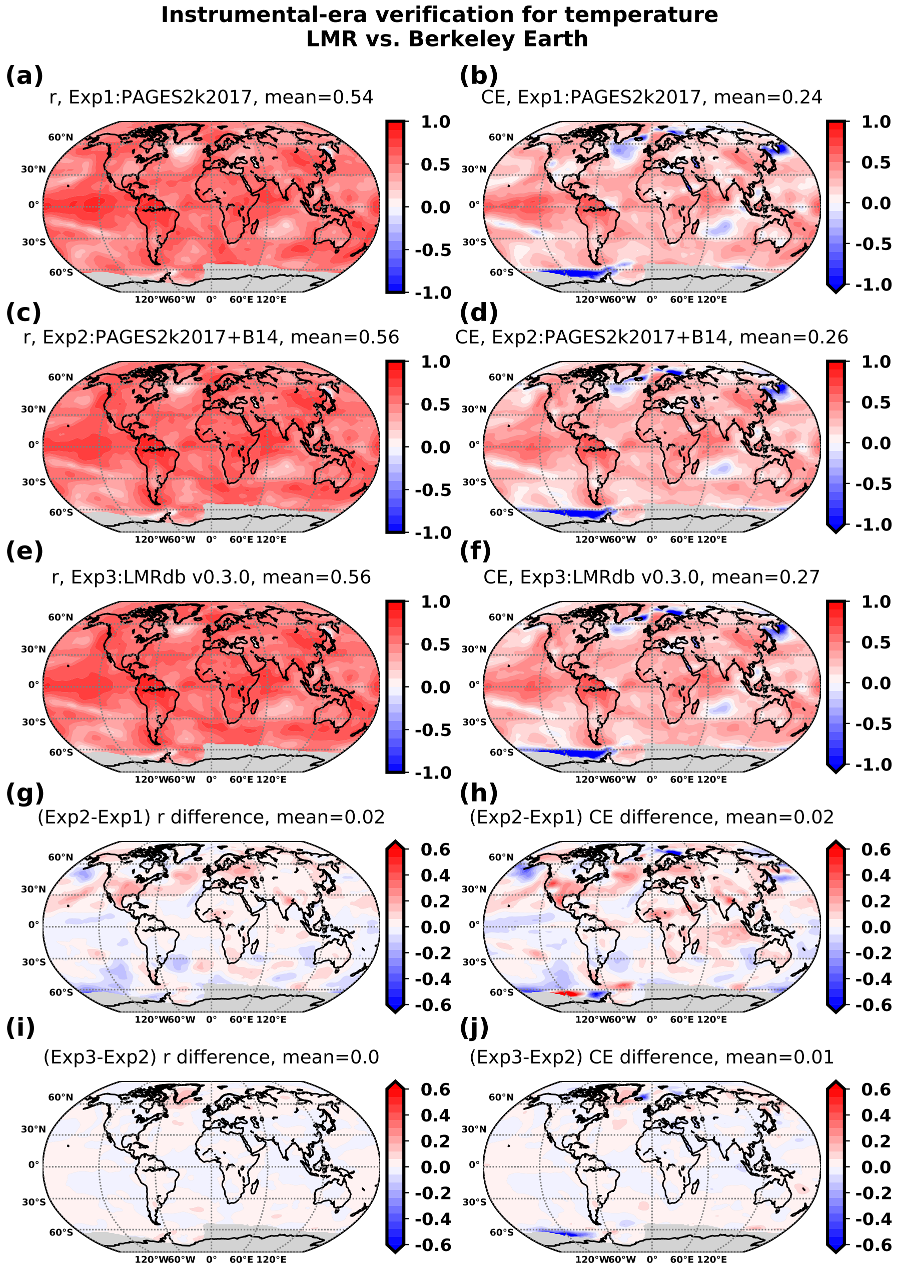

Figure 9. As in Figure 6, but comparing experiments performed with different proxy networks: (a) correlation (r) and (b) CE for experiment 1: PAGES 2k Consortium (2017) proxies only, (c) and (d) experiment 2: with the addition of tree ring chronologies from Breitenmoser et al. (2014), and (e) and (f) experiment 3: with all proxies in the updated LMR database. The differences in correlation and CE between experiments 2 and 1 are shown in (g) and (h) respectively, and between experiments 3 and 2 in (i) and (j). Notice the latter is different to Figure 6, where differences between experiments 3 and 1 are shown.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-120
Manuscript under review for journal Clim. Past
Discussion started: 18 September 2018

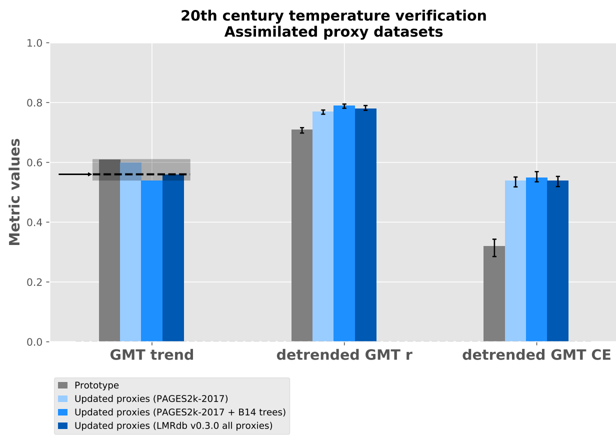

Figure 10. As in Fig. 8 for reconstruction experiments performed with different proxy networks: PAGES 2k Consortium (2017) proxies only, with the addition of tree ring chronologies from Breitenmoser et al. (2014), and with all proxies in the updated LMR database.

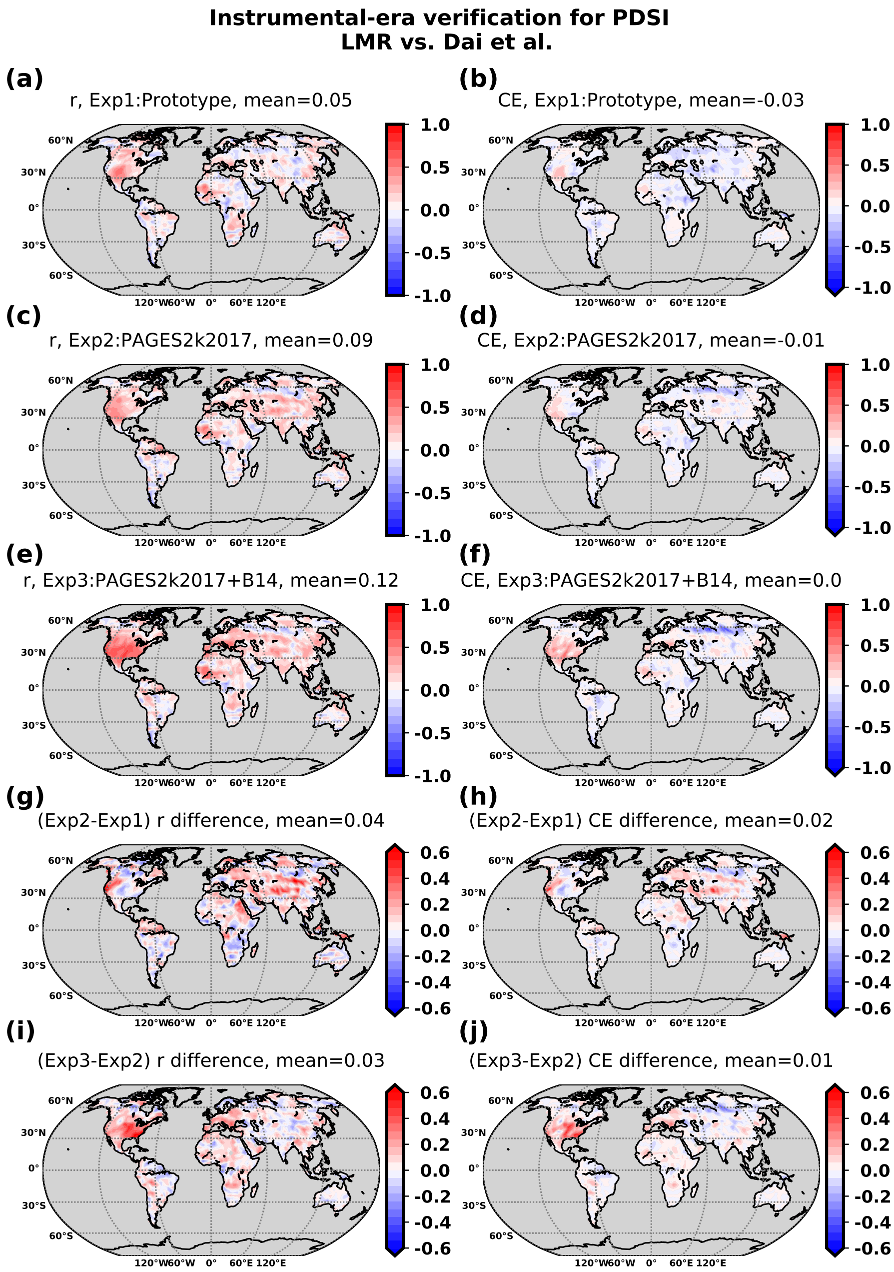

Figure 11. As in Figure 9, but comparing PDSI reconstructions against the Dai (2011) analysis for experiments performed with different proxy networks: (a) correlation and (b) CE for experiment 1: prototype reanalysis already presented in Fig. 4, experiment 2: PAGES 2k Consortium (2017) proxies only, (e) and (f) experiment 3: with the addition of tree ring chronologies from Breitenmoser et al. (2014). The differences in correlation and CE between experiments 2 and 1 are shown in (e) and (f) respectively, and between experiments 3 and 2 in (i) and (j).

Table 1. Summary of instrumental–era verification results for the prototype and updated reanalyses. Verification scores shown are correlation (r) and coefficient of efficiency (CE), for the annual global mean temperature (GMT) and detrended GMT verified against the consensus of instrumental–era analyses, the global mean of gridpoint r and CE characterizing the spatially reconstructed temperature, 500 hPa geopotential height (Z500) and Palmer Drought Severity Index (PDSI). LMR spatial temperature is verified against the Berkeley Earth analysis (Rohde et al., 2013), Z500 is verified against the 20CR-V2 reanalysis (Compo et al., 2011) and PDSI is verified against the Dai (2011) analysis.

Reanalysis	Annual GMT		Detrended GMT		Spatial temperature		Spatial Z500		Spatial PDSI	
	r	CE	r	CE	r	CE	r	CE	r	CE
Prototype	0.91	0.79	0.71	0.32	0.49	0.15	0.41	0.07	0.05	-0.03
Updated	0.93	0.87	0.79	0.60	0.56	0.26	0.46	0.17	0.12	0.01

Table 2. Ensemble calibration ratios characterizing ensembles of global-mean temperature from reconstruction experiments performed with covariance localization applied with various values of the localization radius R_L. The ratio from an experiment without covariance localization is also shown for comparison. The result from the prototype reanalysis is shown for reference.

Localization R_L = 5000 km	Localization R_L = 10000 km	Localization R_L = 15000 km	Localization R_L = 20000 km	Localization R_L= 25000 km	No localization	Prototype (no localization)
0.66	0.50	0.68	0.81	1.06	2.69	0.83

Table B1. Mean differences in Bayesian Information Criterion (ΔBIC) corresponding to PSMs for records within the proxy categories considered in LMR, between models calibrated using proxy seasonal responses from the metadata or derived objectively during calibration, with respect to the reference of annual seasonality. Calibration dataset: GISTEMP v4.

Proxy types	Number of records	Seasonal (metadata)	Seasonal (objective)
Tree ring width (PAGES2k-2017)	347	-1.34	-4.84
Tree ring width (Breitenmoser et al.)	2156	-1.72	-5.24
Tree ring wood density	59	-23.28	n/a
Coral $\delta^{18}O$	75	+0.02	n/a
Coral Sr/Ca	30	-0.01	n/a
Coral Rates	11	+0.03	n/a
Ice core $\delta^{18}O$	89	+0.02	n/a
Ice core δD	12	0.00	n/a
Ice core accumulation	3	0.00	n/a
Ice core melt	1	0.00	n/a
Lake core varve	7	-0.52	n/a
Lake core misc.	2	-2.32	n/a
Bivalve $\delta^{18}O$	1	0.00	n/a
Tree ring $\delta^{18}O$	1	+11.81	n/a

Table B2. Mean differences in Bayesian Information Criterion (ΔBIC) for tree ring width univariate "temperature or moisture" and bivariate PSMs, calibrated using metadata seasonality or derived objectively during calibration, against their respective univariate temperature-only PSMs as reference. Calibration datasets: GISTEMP v4 and GPCC v6.

PSM formulation	Seasonal (metadata)		Seasonal (objective)	
	PAGES 2k trees	Breitenmoser trees	PAGES 2k trees	Breitenmoser trees
Univariate - temperature or moisture	-0.86	-1.41	-2.59	-6.65
Bivariate	+2.63	+1.73	-2.35	-6.88