S1 Introduction

This document provides supplemental information on the characteristics of newly developed statistical proxy system models (PSMs) for tree ring width proxies, and on additional verification results of the updated paleoclimate reconstructions performed within the Last Millennium Reanalysis (LMR) framework.

S2 Proxy seasonal responses

An important update to the LMR proxy modeling capabilities is the introduction of PSMs which include a representation of proxy seasonality. Two methods for defining proxy seasonality are considered: use of the seasonal response information included in the proxy metadata, and objectively determining the seasonal period which leads the best linear fit to the proxy data as part of the PSM calibration procedure. Here we examine the differences between seasonality information from these two approaches. The comparison focuses on PSMs developed for tree ring width (TRW) records as seasonality is particularly important for these proxies (Briffa et al., 2002, 2004). TRW chronologies in our proxy database originate from two distinct data sources, the PAGES 2k Consortium (2017) community-curated collection and records used in Breitenmoser et al. (2014) as processed by Anderson et al. (2018). The metadata describing the records in PAGES 2k Consortium (2017), including seasonality, have been established by community experts, whereas TRW seasonality annotating the Breitenmoser et al. (2014) records was defined using a simple latitude dependency (Anderson et al., 2018). Objective seasonality is described in section 2.4 of the main text.

Figure S1 shows the overall distributions of months defining seasonal temperature responses, compiled across all TRW records and centered on the annual period the proxy data describes. Distributions corresponding to metadata and objectively-derived seasonality information are shown for both TRW data sets. For tree ring records in PAGES 2k Consortium (2017), seasonality metadata information (Figure S1a) shows records are predominantly characterized by a June–August (JJA), i.e. boreal summer, response to temperature, in addition to a significant proportion of trees with an annual seasonality as indicated by the flat distribution outside of the JJA maximum. The distribution describing the Breitenmoser et al. (2014) records (Figure S1c) reflects the simple latitude-dependent approach to define seasonality. Responses are limited to JJA for northern hemispheric (NH) trees and December–February (DJF) for trees located in the southern hemisphere (SH). The handful of trees with an annual response correspond to the few tropical records present in the dataset.

With seasonality information determined objectively during PSM calibration (Figs. S1b and d), a greater diversity in seasonal responses is obtained. This is most striking with the Breitenmoser et al. (2014) chronologies. Responses remain dominated by the NH summer season, however with less emphasis on annual records as evidenced by a slightly increased representation of the boreal spring and austral summer months for NH trees and SH trees respectively. We also note the greater consistency of seasonal responses to temperature between PAGES 2k Consortium (2017) and Breitenmoser et al. (2014) TRW records.

Figure S2 presents the results with respect to temperature and precipitation seasonal responses associated with the bivariate PSMs. Distributions describing temperature seasonality are very similar to those of the univariate models shown in Fig. S1. More importantly, the distributions of the seasonal responses to precipitation shows that objectively-derived responses are significantly different than those from proxy metadata (assumed identical to temperature responses). Distributions show a number of records with sensitivity to precipitation during the nominal growing seasons (boreal summer for NH trees and austral summer for SH trees). A noteworthy feature is the maximum in the distribution shifted toward the boreal winter season in Figs. S2d and h ,consistent with tree ring growth sensitivity to precipitation occurring during the cool season preceding the growth period (St. George et al., 2010; St. George, 2014). Note that the local maximum in the austral winter in Fig. S2d correspond to SH trees. As with the temperature responses, we also note the greater consistency between responses to precipitation for PAGES 2k Consortium (2017) and Breitenmoser et al. (2014) TRW records.

S3 Independent calibration-validation

Verification statistics from a series of reconstruction experiments similar to those presented in section 4 in the main text are presented, but with independent calibration and validation data during the instrumental period. Temperature reconstructions are performed using regression-based PSMs calibrated with data covering the 1920–2015 period, instead of the entire instrumental era (i.e. 1880–2015) as for experiments reported in the main text. To gain a perspective on reconstruction skill which is independent from calibration, verification is performed over the 40-year period of the instrumental era not considered in PSM calibration (i.e. 1880–1919). As in experiments reported in section 4.1 of the paper, only proxies from the PAGES 2k Consortium (2017) data set are assimilated.

A summary of verification skill metrics is shown in Fig. S3. We first look at the trend in the global mean temperature (GMT) characterizing the 40-yr verification period. The GMT trend during the validation period in the instrumental analyses (i.e. GISTEMP, Berkeley Earth, HadCRUT4 and MLOST), is itself found to have a large uncertainty. Values among all products range from slightly positive to about -0.35 K per 100 years (see the gray shading in Fig. S3). Despite this uncertainty, instrumental era observations indicate a cooling trend characterizing the verification period, with a consensus (average of all products) value of -0.2 K per 100 years. Comparing trends in reconstructions generated using the various PSM configurations, we see that all reconstructions are characterized by global cooling, with the trend underestimated with the univariate annual PSMs, and generally overestimated with the seasonal univariate models. The best agreement with the consensus trend is obtained when bilinear seasonal PSMs are used, along with objectively-derived proxy seasonal responses for tree-ring-width proxies. We wish to underline here LMR's ability to recover the GMT cooling during the 1880-1919 period, distinct from the 1920–2015 PSM calibration period, which is characterized by a warming trend. This result supports the fact that reconstructions with LMR are not directly tied to, or limited by, the climate states and their evolution represented

in the calibration data set, especially since the inputs to these PSMs come from climate model priors from a different time period.

With respect to the other metrics considered in evaluating the reconstructed detrended GMT and spatial temperature patterns, the skill over the 40-year verification period is generally less than the entire instrumental era (reported in the main text). A comparison of verification statistics from the 1880–1919 and 1880–2000 periods from the same series of experiments (using PSM calibrated over the complete instrumental-era data) (not shown) confirm that the decrease in skill is a characteristic of the shorter calibration period rather than related to PSMs calibrated over a particular subset of the available calibration data.

Comparing verification metrics for different PSM configurations, skill in the detrended GMT is maximized for PSMs formulated with seasonal responses from proxy metadata, particularly when tree ring width proxies are modeled with a bilinear formulation on temperature and moisture. The skill from PSMs based on objectively-derived seasonal responses are generally lower, but remains similar to results obtained with other PSMs. Skill in the representation of spatial patterns, as summarized with the global mean of the CE values calculated at every reanalysis grid points, is less skillful in reconstructions using annual univariate PSMs. The most skillful reconstruction, as indicated by the least negative mean of the gridded CE, is obtained with seasonal objective PSMs, with a bilinear formulation to model tree ring width proxies.

The independent calibration–validation results reported here are therefore generally consistent with the findings presented in the main text. In particular, the least accurate reconstructions are obtained with univariate annual PSMs, whereas reconstructions using seasonal PSMs with objectively-derived seasonal responses, along with a bivariate temperature and moisture formulation for tree ring width proxies, are found to be more skillful. Contrasts in skill are notable for GMT trends and the representation of temperature spatial patterns. Therefore, despite the decrease in the robustness of statistics inherent to verification performed over a shorter time period, the independent calibration–validation experiments support the selection of the PSM configuration reported in section 3 of the main text.

S4 Proxy-space verification

The verification results discussed above and in the main text were based on instrumental data covering the modern era. An evaluation of paleoclimate reanalyses over a period prior to the instrumental era is essential for assessing reanalysis performance in the absence of calibration constraints. Here we use proxy-based verification, consisting of a comparison of proxy time series estimated (forward-modeled) from the posterior (i.e. the reconstructions) and the actual proxy observations. Verification of proxy estimates obtained from the uninformed climate-model prior used in LMR serve as a reference for comparison.

The coefficient of efficiency (CE) (Nash and Sutcliffe, 1970) is used as the verification skill metric. Specifically, we use the change in CE (ΔCE) between the posterior proxy estimates and estimates obtained from the prior, $\Delta\text{CE} = \text{CE}_{posterior} - \text{CE}_{prior}$. This represents a measure of the information (in proxy-space) resulting from the assimilation. A ΔCE value is

obtained for every proxy record considered (assimilated and withheld from assimilation), and two summary measures are considered: the fraction of proxy records which are characterized by a positive ΔCE (i.e. proxy records more accurately represented in the posterior than in the prior), and the mean across all proxy time series of the ΔCE. The latter provides a summary measure of how the reanalysis is more skillful than the prior, as represented in proxy space. Summary skill metrics are compiled over two distinct time intervals: 1880–2000CE, i.e. period over which PSMs have been calibrated, and 1–1879CE i.e. prior to the calibration period. Results are shown in Fig. S4.

This proxy-based verification indicates the majority of proxy records are characterized by an improved representation in the reanalysis compared to the prior, including proxies that are withheld from assimilation. The mean ΔCE are also found to be positive, with values ranging from +0.15 to +0.45. The fraction of records with an increase in CE is larger for assimilated proxy records compared to withheld proxies, regardless of the verification period (i.e. within or outside the calibration period). The mean CE increase is also generally smaller for withheld records. This is not surprising, as the improvement to the fit of the non-assimilated proxies rely solely on the remote information provided by the other (i.e. assimilated) records through spatial ensemble-estimated covariances. Despite the smaller changes in CE for non-assimilated proxies, the majority of records remain characterized by a improved fit in the reanalysis as indicated by a positive mean ΔCE across records.

Focusing on the verification statistics from reconstructions using the different PSM configurations, we find that the least accurate reconstructions are obtained with univariate PSMs calibrated on annual temperatures, confirming the verification results based on instrumental products. The fraction of assimilated records with a positive change in CE, and the mean of ΔCE, both increase for reconstructions using seasonal PSMs. Further improvement is obtained with bivariate PSMs to forward-model tree ring width proxies. These findings characterize both the in- and out-of-calibration periods. The improved fits in the posterior are indicative of increased weighing of these proxies in the reanalysis resulting from reduced observation error variances, estimated from the residuals characterizing the regression-based PSMs with respect to calibration data as described in the main text. These results are also indicative of a preserved consistency with other nearby assimilated proxies influencing the reanalysis at the location of the proxy record.

For withheld proxies, differences in ΔCE do not differentiate between experiments. The proportion of non-assimilated records with a positive ΔCE is roughly similar among all experiments, with only a slight advantage for seasonal (objectively-determined responses) univariate PSMs. The mean ΔCE is also equivalent across experiments (at the 95% confidence level).

Proxy-space verification provides an independent alternative to the evaluation of reconstructions based on comparisons with available instrumental data analyses. We find here that results are generally consistent with instrumental data verification, in that seasonal PSMs lead to improved representations of proxy observations in the reanalysis for the assimilated records, compared to the annual univariate models on temperature used in the LMR prototype; however, verification statistics on the withheld proxies are not as conclusive.

S5 Covariance localization

As described in the main text, a series of climate reconstruction experiments are conducted to assess the impact of applying covariance localization in the data assimilation algorithm; here we use the Gaspari-Cohn localization function (Gaspari and Cohn, 1999). Reconstruction experiments consider all proxy records in the expanded database. In the main text, we focused on the relationship between the mean error in the reanalysis and ensemble spread (i.e. ensemble calibration). Here we complement these results by examining temperature reconstruction skill in terms of the correlation and CE skill metrics to against instrumental-era temperature analyses.

We considered covariance localization radii in the range of 5000–25000 km (Fig. S5), alongside results from an experiment performed without localization. Among the most sensitive metrics to covariance localization are the 20th century GMT trend, CE of the detrended GMT, and the global-mean value of the spatial CE skill scores. The latter corresponds to CE values at every reanalysis grid point obtained from comparisons with the four instrumental data analyses and spatially averaged. We find that 20th century GMT trend is underestimated for small localization radii, due to the limited ability of proxy records to inform the global-mean. The most accurate representation of the GMT trend is obtained when covariances are not localized at all. However, localization with the largest radius considered here has a GMT trend underestimated by only about 10%. For the detrended GMT, both the correlation and CE are maximized for localization radii in the 20000-25000 km range, in contrast to a significantly lower CE for this measure when localization is not applied. This suggests that the spatial influences of some noisy proxy records are not sufficiently mitigated, with an adverse impact on the representation of GMT interannual variability. The spatial CE skill metric is also particularly sensitive to the use of covariance localization in the reanalysis, with the least skillful reconstructions obtained with the shorter localization radii. The highest skill is obtained without localization. However, covariances localized with a 25000 km cut-off length scale lead to a mean spatial CE decreased by about 4%. Moreover, the annual spatial anomaly correlations, averaged over all years in the instrumental era and over verification results compiled over the four instrumental verification datasets show a slightly higher value corresponding to a 25000 km localization radius compared to the no localization result.

Therefore, the consideration of ensemble characteristics (i.e. ensemble calibration) as discussed in the main text, as well as trade-offs between the representation of the GMT trend and interannual variability, and the representation of spatial patterns in reanalysis fields, we conclude that covariance localization with a cut-off radius of 25000 km performs best overall.

References

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 multiproxy database (LMRdb v1.0.0), Data Science Journal, submitted, 2018.

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., Schweingruber, F. H., Jones, P. D., Shiyatov, S. G., and Vaganov, E. A.: Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals, Holocene, 12, 737–757, https://doi.org/10.1191/0959683602hl587rp, 2002.

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.

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.

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.

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.

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.

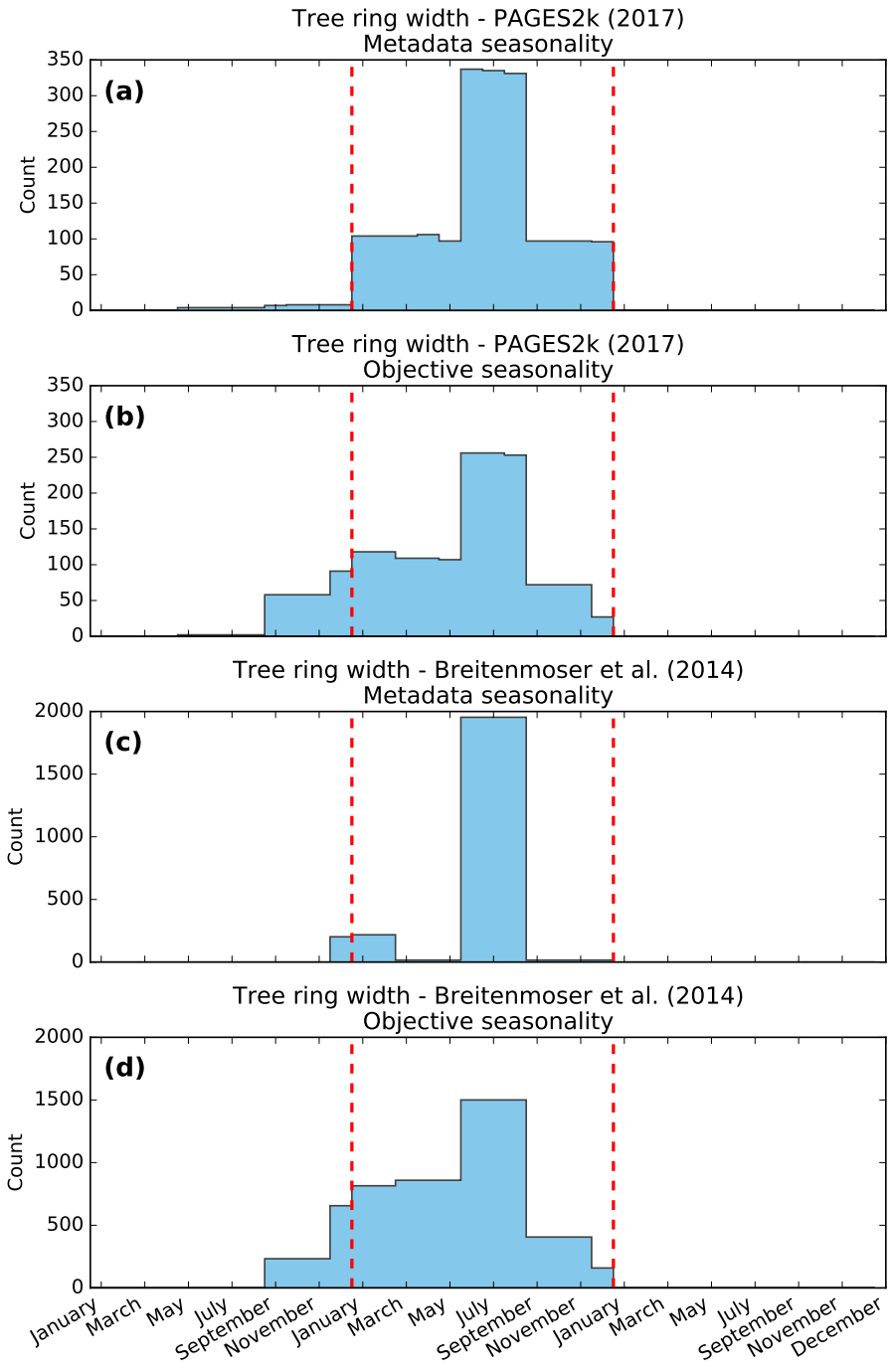

Figure S1: Distributions of months of the year included in proxy seasonal responses for tree ring records in (a) and (b) PAGES 2k Consortium (2017), (c) and (d) Breitenmoser et al. (2014) data sets. For each proxy set, histograms are shown describing the seasonality information contained in (a) and (c) the proxy metadata, and (b) and (d) objectively-derived during PSM calibration using a goodness-of-fit approach. Vertical dashed red lines delineate the annual period of the modeled proxy. PSM calibration is performed with respect to temperature, using the GISTEMP v4 dataset.

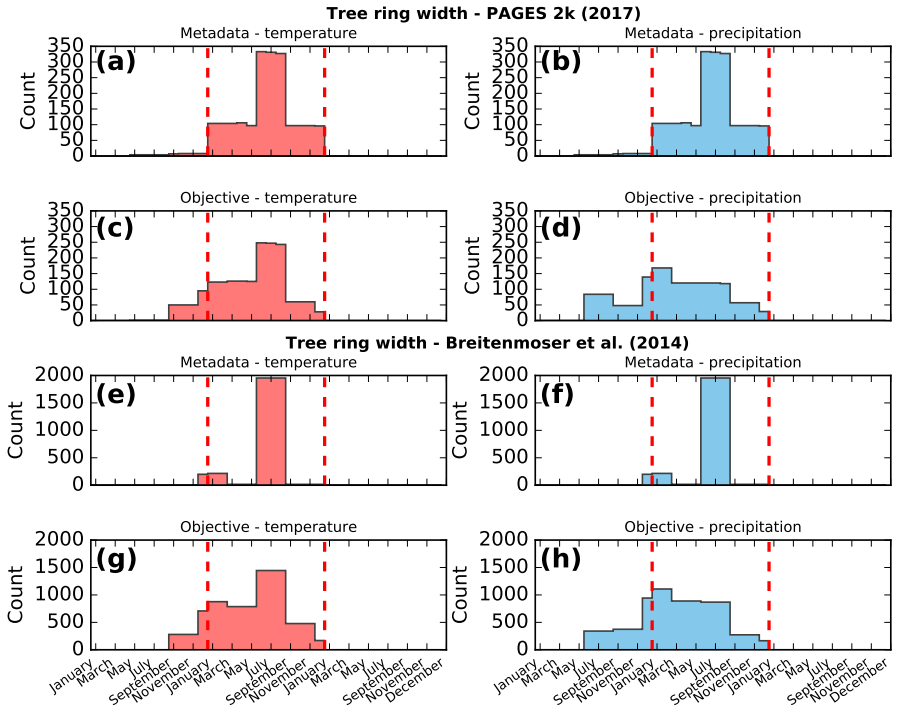

Figure S2: As in Fig. S1, but for seasonal responses of temperature and precipitation used in the bivariate TRW PSMs, for tree ring records in (a)–(d) PAGES 2k Consortium (2017), (e)–(h) Breitenmoser et al. (2014) data sets. For each proxy set, histograms are shown describing the seasonality information contained in (a), (b), (e) and (f) the proxy metadata, and (c), (d), (g) and (h) objectively-derived using a goodness-of-fit approach during PSM calibration. Distributions for temperature are shown in the left panels, and for precipitation in the right panels. Bivariate PSM calibration is performed with respect to temperature and precipitation, using the GISTEMP v4 and GPCC v6 datasets.

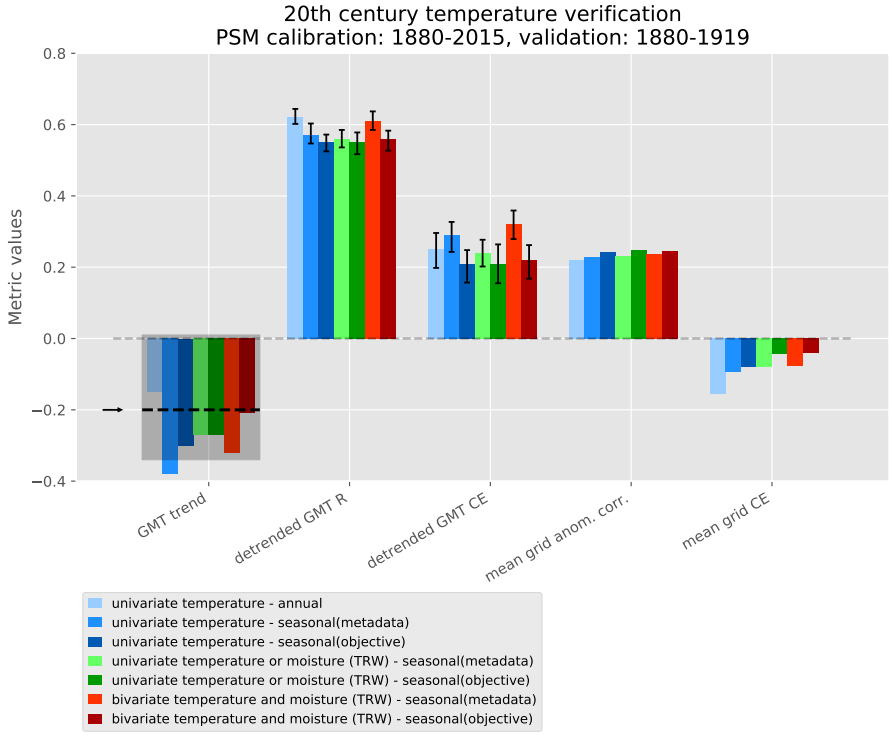

Figure S3: Summary skill metrics of LMR temperature reconstructions against the consensus of instrumental era temperature data sets (GISTEMP, MLOST, Berkeley Earth, HadCRUT4) over the 1880–1919 period. Reconstruction experiments are performed with PSMs calibrated on data covering the 1920–2015 period (excluding the verification period) with the following PSM configurations: univariate on annual temperature (light blue), univariate on seasonal (metadata) temperature (blue), univariate seasonal models with objectively-derived seasonality (dark blue), seasonal (metadata) univariate "temperature or moisture" for tree ring widths and temperature only for all other proxies (light green), same as previous but with objectively-derived seasonality (dark green), bivariate seasonal (metadata) models for tree ring widths and univariate on temperature for all other proxies (red) and same as previous but with objectively-derived seasonality (dark red). Metrics shown are the 20th century trend in the global mean temperature (GMT), correlation and coefficient of efficiency (CE) for the detrended GMT, mean of anomaly correlations against the instrumental data sets, and global mean of gridpoint CE averaged across the same verification data sets. 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.

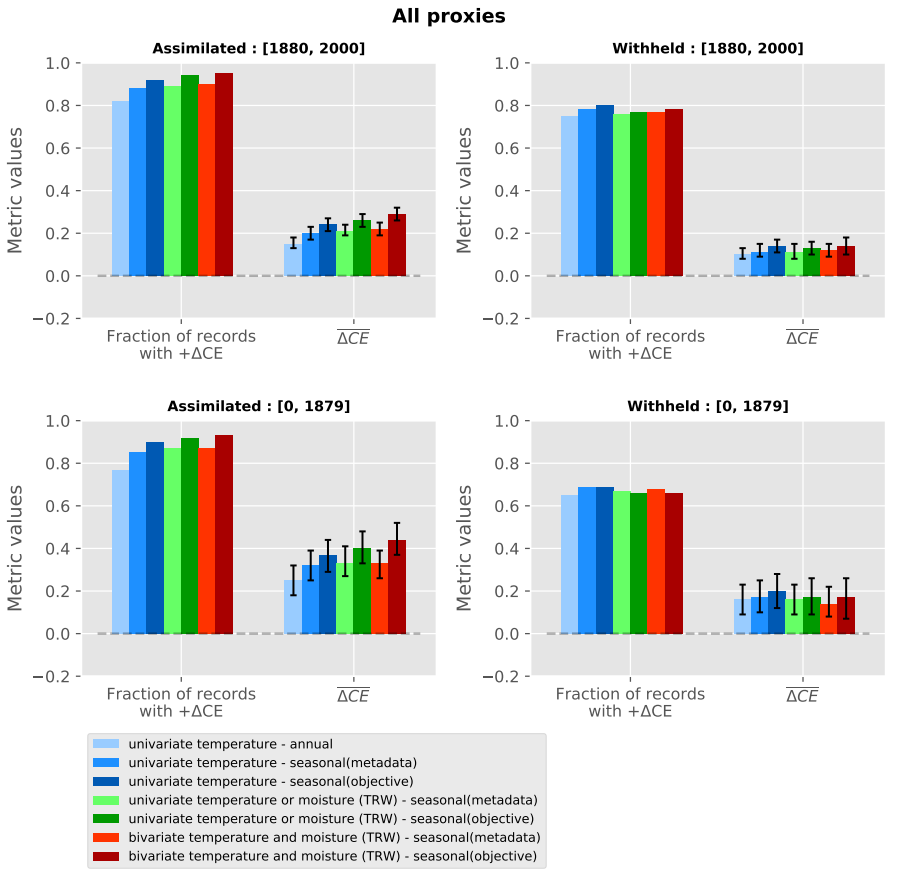

Figure S4: Verification in proxy space of climate reconstructions performed with various PSM configurations. Metrics shown are the proportion of proxy records for which the difference in coefficient of efficiency (CE), evaluated in proxy space, between the prior and posterior (i.e. reanalysis) is positive, and the mean of this difference in CE across all proxy records. Results are shown for assimilated and withheld proxies, for two distinct periods: 1880–2000 (calibration period) and 1–1879 (prior to the calibration period). Error bars are the 5-95% bootstrap confidence intervals on the corresponding skill metric.

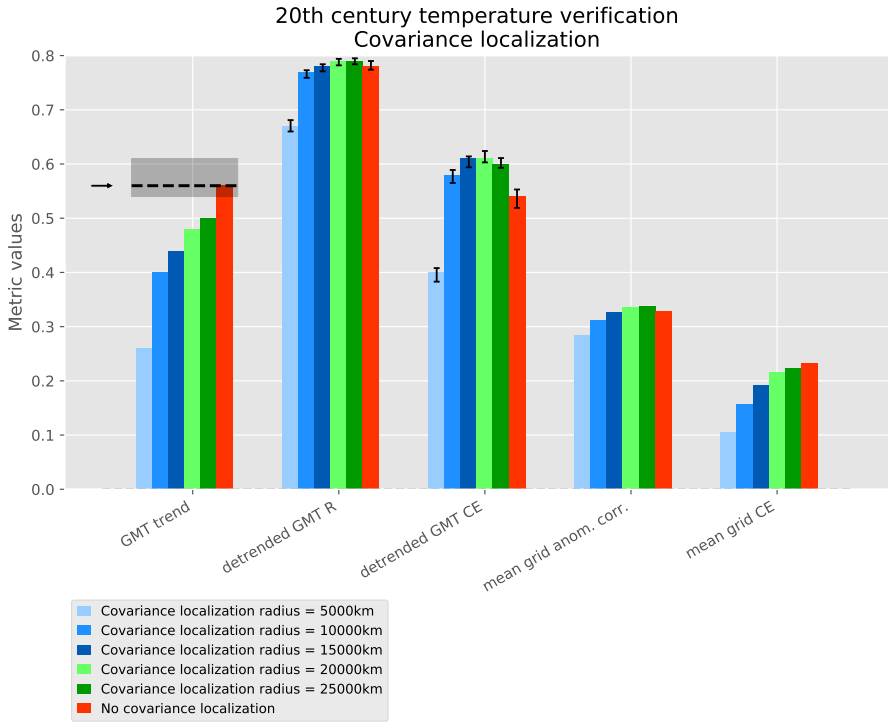

Figure S5: As in Fig. S3 for reconstruction experiments performed with covariance localization, with range of values for the cutoff radius. Verification is performed over the entire 1880–2000 period. The results for an experiment performed without covariance localization is shown for reference.