

Estimating Central Equatorial Pacific SST Variability over the Past Millennium. Part I: Methodology and Validation

JULIEN EMILE-GEAY

Department of Earth Sciences, University of Southern California, Los Angeles, California

KIMBERLY M. COBB

Department of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia

MICHAEL E. MANN

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

ANDREW T. WITTENBERG

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

(Manuscript received 8 September 2011, in final form 5 November 2012)

ABSTRACT

Constraining the low-frequency (LF) behavior of general circulation models (GCMs) requires reliable observational estimates of LF variability. This two-part paper presents multiproxy reconstructions of Niño-3.4 sea surface temperature over the last millennium, applying two techniques [composite plus scale (CPS) and hybrid regularized expectation maximization (RegEM) truncated total least squares (TTLS)] to a network of tropical, high-resolution proxy records. This first part presents the data and methodology before evaluating their predictive skill using frozen network analysis (FNA) and pseudoproxy experiments. The FNA results suggest that about half of the Niño-3.4 variance can be reconstructed back to A.D. 1000, but they show little LF skill during certain intervals. More variance can be reconstructed in the interannual band where climate signals are strongest, but this band is affected by dating uncertainties (which are not formally addressed here). The CPS reliably estimates interannual variability, while LF fluctuations are more faithfully reconstructed with RegEM, albeit with inevitable variance loss. The RegEM approach is also tested on representative pseudoproxy networks derived from two millennium-long integrations of a coupled GCM. The pseudoproxy study confirms that reconstruction skill is significant in both the interannual and LF bands, provided that sufficient variance is exhibited in the target Niño-3.4 index. It also suggests that FNA severely underestimates LF skill, even when LF variability is strong, resulting in overly pessimistic performance assessments. The centennial-scale variance of the historical Niño-3.4 index falls somewhere between the two model simulations, suggesting that the network and methodology presented here would be able to capture the leading LF variations in Niño-3.4 for much of the past millennium, with the caveats noted above.

1. Introduction

The tropical Pacific Ocean is the stage for powerful air–sea interactions, whose oceanic expression ripples through the climate system on a variety of time scales

(e.g., Dijkstra and Neelin 1995; Cane 2005; Emile-Geay 2006). The epitome of such interactions is the El Niño–Southern Oscillation (ENSO) in the interannual range, although variability exists on decadal (Graham 1994), interdecadal (Deser et al. 2004), centennial (Cobb et al. 2003), and longer time scales (Tudhope et al. 2001). The nature of future tropical Pacific trends in sea surface temperature (SST) will shape the response of the large-scale atmospheric circulation to global warming, with profound implications for regional hydroclimate regimes (Seager et al. 2009; Seager and Vecchi 2010).

Corresponding author address: Julien Emile-Geay, Department of Earth Sciences and Center for Applied Mathematical Sciences, University of Southern California, 3651 Trousdale Parkway, Los Angeles, CA 90089-0740.
E-mail: julieneg@usc.edu

Despite its importance, the evolution of tropical Pacific SST under continued anthropogenic forcing is still subject to considerable uncertainties (Meehl et al. 2007; DiNezio et al. 2009; Collins et al. 2010; Vecchi and Wittenberg 2010). Constraining the tropical Pacific's sensitivity to greenhouse forcing requires knowledge of its response to natural (solar and volcanic) radiative perturbations on decadal to centennial time scales. This type of problem is well posed for general circulation models (GCMs; Meehl et al. 2003), provided that their output can be validated against some estimate of natural climate variability. Because nonlinearities in the climate system complicate the separation of responses to natural and anthropogenic forcing, estimating the envelope of natural climate variability compels the investigator to look beyond the onset of the anthropogenic era, which corresponds to the best-observed interval in Earth's history. In the tropical Pacific, the sparse network of instrumental climate observations rarely extends before 1900 and is marked by considerable discrepancies between SST datasets (Vecchi et al. 2008; Deser et al. 2010). Adding to this challenge, there exist very few long, high-resolution paleoclimate records from the area. Published records from the circum-Pacific region paint a coarse and heterogeneous picture of its climate of the past millennium (Moy et al. 2002; Graham et al. 2007; Conroy et al. 2009; Sachs et al. 2009; Tierney et al. 2010), illustrating the difficulty of extrapolating large-scale climate patterns from a select handful of individual proxy records, and highlighting the need for a unifying mathematical framework in which all the relevant proxy data can be synthesized.

Multiproxy reconstructions of tropical Pacific SST aim to combine the common signals contained in a comprehensive network of proxy records to provide continuous, quantitative reconstructions that can be rigorously compared against radiative forcing histories and climate model output. Despite several papers published to date, there is little consistency among reconstructions (Gergis et al. 2006). As recently reviewed by Wilson et al. (2010, hereafter W10), the earliest attempt combined a dozen annually resolved ENSO-sensitive proxy records, mostly corals and tree rings, to reconstruct ENSO variability over the last 300 years (Stahle et al. 1998). Using a more comprehensive network of 112 annually resolved proxy records (both tropical and extratropical), Mann et al. (2000) produced Niño-3¹ reconstructions for the past four centuries, quite highly correlated to that of Stahle et al. (1998) over their

period of overlap, as expected from the number of common series (Adams et al. 2003). Evans et al. (2002) used a network of tropical Pacific coral records to reconstruct the first two modes of tropical Pacific SST, with limited reliability prior to A.D. 1800 or so. Other attempts focused on interannual oscillations obtained from long, replicated tree-ring chronologies (D'Arrigo et al. 2005) and other records (Braganza et al. 2009; Gergis et al. 2006; Gergis and Fowler 2009), in lieu of reconstructing low-frequency (LF) variations. W10 expand on previous work by considering several carefully chosen networks of high-resolution tropical records to reconstruct ENSO using three statistical methods. However, they explicitly filtered out LF signals from corals to focus on interannual variations, representing a limitation with regard to the assessment of LF variability. Their reconstructions stopped approximately A.D. 1600, but they report little consistency among the three methodologies they implemented before A.D. 1800 or so, even with identical proxy networks. This highlights the need for improved statistical methodologies and/or more proxies located from the central tropical Pacific to extract robust signals from such networks.

Using a global multiproxy database and a climate field reconstruction technique, Mann et al. (2009, hereafter M09) were able to extend a near-global surface temperature field back to A.D. 500, yielding the most comprehensive picture to date of the climate of the past millennium. In the tropical Pacific, they found the Niño-3 index to exhibit a slow, millennial-scale warming trend characterized by a relative cooling during the Medieval Climate Anomaly (MCA) compared to the Little Ice Age (LIA), which was absent from the numerical simulations performed with two coupled GCMs [Goddard Institute for Space Studies Model E-R (GISS-ER) and the National Center for Atmospheric Research (NCAR) Climate System Model, version 1.4 (CSM1.4)]. Because the method was global and contained a modest number of records from the tropical Pacific, the Niño-3 signal depends substantially on large-scale teleconnections that, in principle, might not be stationary. Its strict reliance on continuous records also precludes the use of some valuable ENSO records (e.g., Cobb et al. 2003). Indeed, most efforts to date have focused on annually resolved, continuous proxy records, resulting in a bias toward tree rings from the extratropics.

In this study, we use a complementary approach to reconstruct tropical Pacific SST variability over the past millennium. By focusing on tropical records from circum-Pacific regions, and by relaxing the requirement of continuous, annually resolved proxies, we explicitly target LF tropical Pacific variability, as well as intermittent interannual signals over the past millennium.

¹ Defined as the average SST anomaly in the region 5°N–5°S, 150°–90°W.

We apply the same method as Mann et al. (2008)—hybrid regularized expectation maximization (RegEM) truncated total least squares (TTLS), described in section 3b—to reconstruct a widely used ENSO index on annual to multicentennial time scales. This differs from the M09 methodology in that we target a single index, not a climate field. The focus on a single index, although eliminating spatial information, yields a simpler estimate that better lends itself to sensitivity analyses. It also allows the simpler composite-plus-scale (CPS) method (Bradley and Jones 1993; Jones and Mann 2004) to be used as a point of comparison. In the first part of this two-part paper, we extensively evaluate both methods using frozen network and pseudoproxy tests, assessing the reliability of the reconstructed indices. In the second part (Emile-Geay et al. 2013, hereafter Part II), we apply these methods to extend three instrumental datasets [Extended Reconstructed SST, version 3 (ERSSTv3), the Second Hadley Centre SST dataset (HadSST2), and Kaplan SST] over the past millennium. It will be shown that, to some extent, details of the statistical methodology are secondary, and that much of the uncertainty between these preinstrumental tropical SST reconstructions stems from discrepancies among the historical SST analysis products themselves. Nonetheless, they enable robust conclusions that challenge our current understanding of LF tropical Pacific variability, with important implications for climate prediction beyond the interannual scale.

The paper is structured as follows: we start with a description of the proxy database (section 2) and statistical methodology (section 3); we then validate both methods over the historical interval via frozen network analysis (section 4) and pseudoproxy experiments (section 5). Results are discussed in section 6, providing an outlook toward Part II, in which several reconstructions of Niño-3.4 are analyzed and implications for GCM representations of the tropical Pacific climate response to radiative forcing are discussed.

2. Data

a. Target temperature data

To alleviate computational complexity, we focus on a simple measure of ENSO (the Niño-3.4 index, defined as the average SST anomaly in the region 5°N–5°S, 170°–120°W; e.g., Trenberth 1997). The Niño-3.4 index is closely related to the first principal component (PC1) of monthly SST in all three major historical SST analyses: HadSST2 (Rayner et al. 2006), Kaplan SST (Kaplan et al. 1998), and ERSSTv3 (Smith et al. 2008). Both Kaplan and ERSSTv3 use some form of interpolation to

impute missing values when they are missing. Because HadSST2 does not, we performed the imputation via the RegEM algorithm (Schneider 2001) with individual ridge regressions.² We refer to the resulting dataset as HadSST2i. Niño-3.4 indices calculated from the three datasets are plotted in Fig. 1.

Since ENSO reaches its peak amplitude and displays most pronounced teleconnections during Northern Hemisphere winter (Bjerknes 1969; Horel and Wallace 1981; Ropelewski and Halpert 1987), we annualized datasets by computing December–February (DJF) averages, where it is understood that January and February belong to year 0 while December belongs to year –1, in the terminology of Rasmussen and Carpenter (1982). To account for the strong decrease in proxy availability after 1995, our historical period ends that year. Although the three datasets share a large fraction of measurements [from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS); Worley et al. 2005], it is clear that the various quality-control and interpolation strategies lead to significant discrepancies before approximately A.D. 1950, with ERSSTv3 displaying notably less variance compared to the other two datasets, especially before 1877. This is at odds with documentary accounts (Quinn 1993; Ortlieb and Macharé 1993; Whetton and Rutherford 1994; DeVries et al. 1997; Davis 2001) and annually banded coral data (Urban et al. 2000; Evans et al. 2000; Ault et al. 2009), which depict intense ENSO activity during that period. The Kaplan SST analysis, on the other hand, was designed to retain variance of the leading (large scale) eigenmodes, which sometimes requires overwriting data entries (Kaplan et al. 2003), and it exhibits the largest decadal variability before approximately 1875. However, Kaplan et al. (1997) warn that a caveat of their methodology is to artificially redden the resulting SST field; that is, increasing LF variability at the expense of high-frequency (HF) variability. The HadSST2i data go back farthest in time (A.D. 1850), albeit with large amounts of missing values before approximately A.D. 1950, whose imputation did not overwrite existing data. Its late-nineteenth-century variability falls between Kaplan and ERSSTv3 (Fig. 1a).

A look at LF instrumental variability (Fig. 1b) reveals large discrepancies among Niño-3.4 trends, whose magnitudes vary by more than a factor of 2: the smallest trend is found for Kaplan SST ($0.18^{\circ}\text{C century}^{-1}$),

² Ridge regression with a data-adaptive choice of regularization parameter was shown by Schneider (2001) to outperform truncated total least squares for monthly SST imputations. Such is not the case with paleoclimate reconstructions, as we shall see shortly.

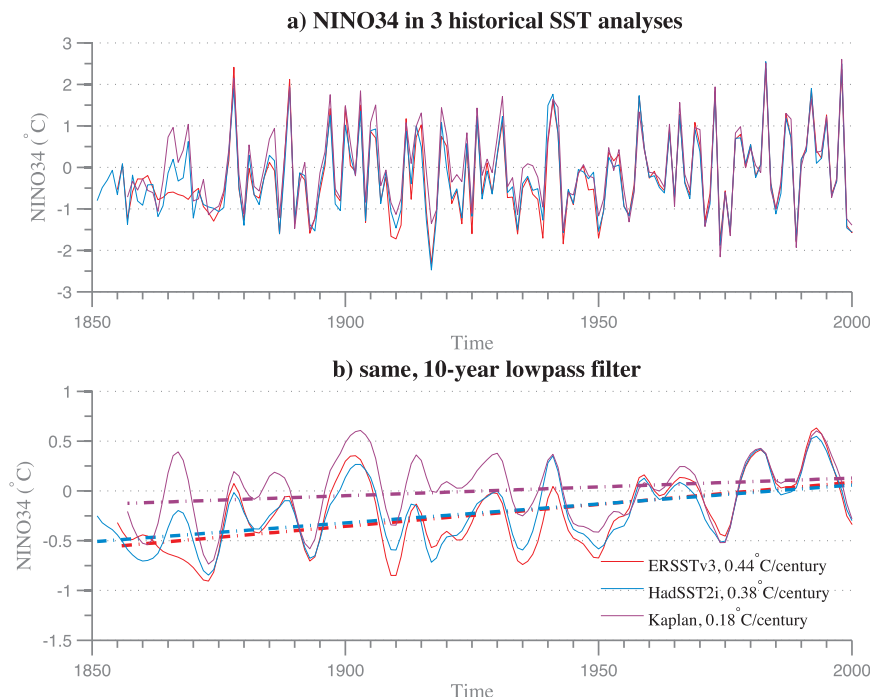


FIG. 1. (a) Niño-3.4 index in three major datasets: ERSSTv3 (Smith et al. 2008; red), HadSST2i (Rayner et al. 2006; blue), and Kaplan SST (Kaplan et al. 1998; purple). Results shown are DJF averages of monthly data. (b) The 10-yr low-pass-filtered time series; linear trends are shown in dashed lines of the respective color, with magnitudes given in the figure legend.

followed by HadSST2i ($0.38^{\circ}\text{C century}^{-1}$) and ERSSTv3 ($0.44^{\circ}\text{C century}^{-1}$). These discrepancies are well documented (Vecchi et al. 2008; Karnauskas et al. 2009) and are known to be principally due to differing interpolation methods, treatments of satellite-derived data, and instrumental bias adjustments (Deser et al. 2010; Yasunaka and Hanawa 2010). Such discrepancies between instrumental SST data products will prove to be crucial sources of uncertainty in our estimates of Niño-3.4 variability before the historical era. However, whenever illustration is warranted we use HadSST2i as our target, as a middle-of-the-road estimate of the historical Niño-3.4 trajectory.

b. Proxy network

Our network gathers paleoclimate archives from ENSO-teleconnected regions (Horel and Wallace 1981; Ropelewski and Halpert 1987; Trenberth et al. 1998) using the following criteria: (i) located equatorward of 35° , (ii) temporal resolution of 5 yr or finer, (iii) absolute dating uncertainties of less than ± 5 yr, (iv) documented sensitivity to surface temperature and/or precipitation, and (v) availability before A.D. 1851 and after A.D. 1974 (though not necessarily continuously). These criteria being arguably subjective, we screened the network for

significant correlation with the raw Niño-3.4 index (section 3a), providing an objective measure of a proxy series' relevance to a tropical Pacific climate reconstruction. (The full proxy database is available online at http://hurricane.ncdc.noaa.gov/pls/paleox/f?p=519:1:4253936273493430:::P1_STUDY_ID:13684 and is catalogued in the appendix in Table A1, with boldface indicating those records that are significantly correlated with the HadSST2i Niño-3.4 index.) All records are publicly available from the National Climatic Data Center (NCDC), with one exception (updated Palmyra record; K. M. Cobb et al. 2013, unpublished manuscript). The results of the screening differ slightly for each of the three target SST datasets used (by at most one series retained).

Our database differs significantly from that constructed by M09 in pursuit of global temperature reconstructions and W10 in pursuit of multiproxy reconstructions of tropical Pacific climate. M09 consider records that are continuous and without latitudinal limits; some sedimentary or speleothem records have resolutions lower than 5 yr. In contrast, the W10 study uses continuous, annually or better-resolved tropical records (corals, tree rings, and ice cores) to reconstruct Niño-3.4 since A.D. 1650. In loosening the requirement for continuous, annually resolved tropical proxy data,

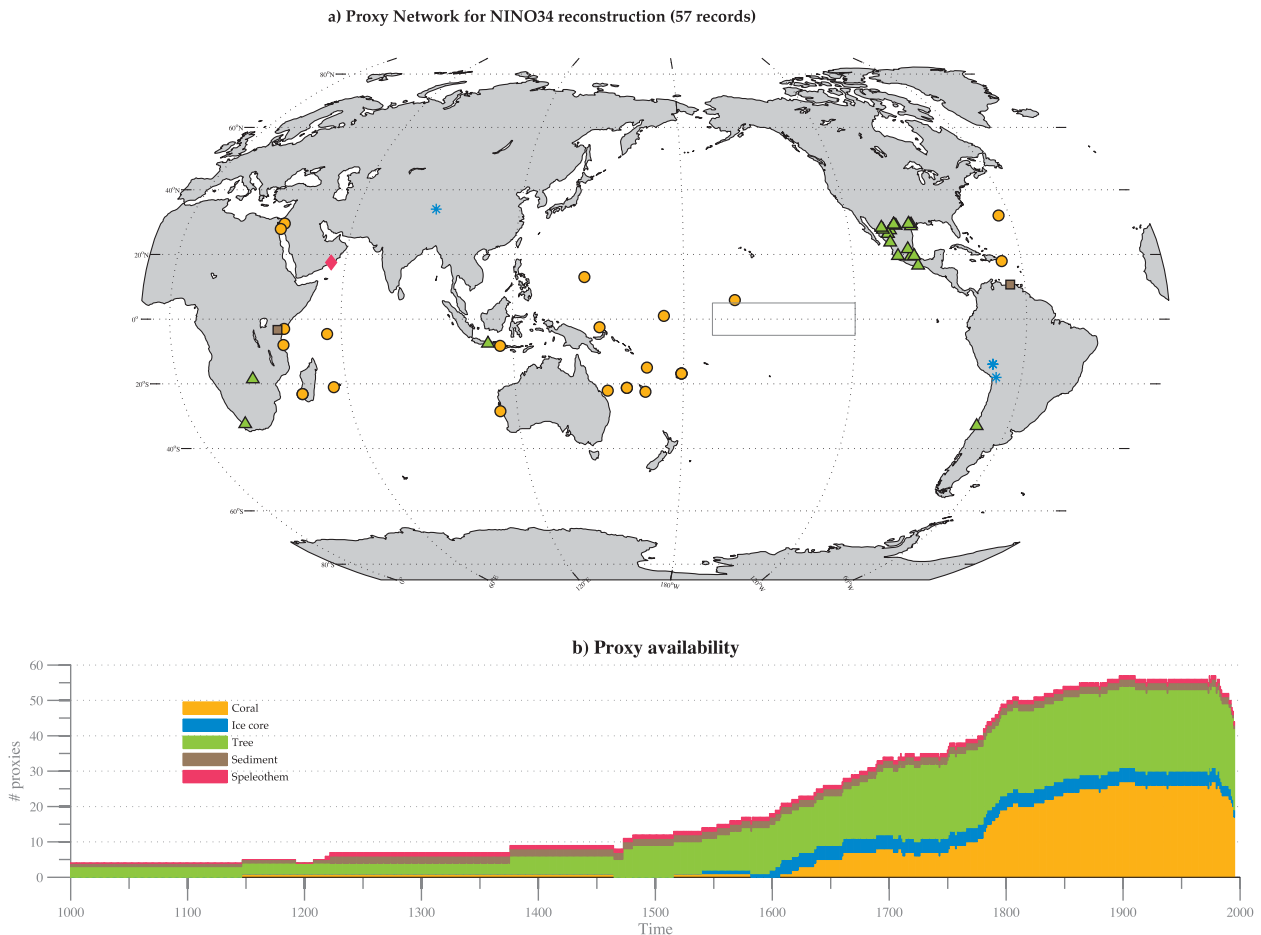


FIG. 2. Proxy database of spatiotemporal characteristics. (a) Location of proxy archives. The Niño-3.4 region is indicated by the black box. (b) Temporal distribution of proxies.

we seek to provide a more accurate view of decadal- to centennial-scale tropical Pacific climate variability back through the MCA. Ours is also the first ENSO reconstruction effort to incorporate discontinuous series (the Palmyra and Mafia Island coral and Lake Challa records) and comprises the latest available data.

Figure 2a displays the geographical distribution of the proxy database by proxy type. Coral sites are distributed throughout tropical oceans, mainly the Pacific. Tree-ring records are mostly confined to areas outside the tropics where seasonality is large enough for visible annual rings to enable construction of absolute annual chronologies. Speleothems are available from caves in several tropical sites, but none of these records passed the screening test (section 3a). Tropical ice cores and lake records are located farthest from the ENSO center of action, and are therefore more vulnerable to shifting teleconnections and/or climate variability not endogenous to the tropical Pacific. Only one marine sediment core (Black et al. 2007) contains sufficiently high-resolution data to be included

in our database. After screening for significant Niño-3.4 correlations, only one ice core record (Quelccaya; Thompson et al. 2006a) is included in the screened proxy network, which is therefore mostly composed of coral and tree-ring records (all annually resolved).

The temporal coverage of our proxy network is displayed in Fig. 2b. The reader will note the steep decline in the availability of coral records over time, with virtually no series available prior to A.D. 1650, with the notable exception of the Palmyra fossil coral chronology (Cobb et al. 2003), which extends back to A.D. 932, albeit discontinuously, covering about 600 of the past 1000 years (K. M. Cobb et al. 2013, unpublished manuscript).

3. Methodology

a. Preprocessing

All series were preprocessed according to the following protocol:

- **Annualization:** Series with subannual resolution were annualized by extracting the December–March months, corresponding to the season of most pronounced ENSO teleconnections. Proxies with a time resolution coarser than one year were linearly interpolated to an annual time scale.
- **Filtering:** We applied a spline filter (Cook and Peters 1981; Weinert 2009) with a 10-yr low-pass cutoff to generate LF (periods longer than 10 yr) and HF (periods shorter than 10 yr) proxy networks.
- **Completeness:** Only proxies available before A.D. 1850 and at least until A.D. 1975 were included, to ensure a long enough calibration.
- **Power transform:** Those series that deviated from Gaussianity were subjected to a power transform (Box and Cox 1964), a classical way to render them approximately Gaussian so they can be incorporated into this parametric statistical framework.
- **Screening:** Following Mann et al. (2008), we performed reconstructions on both unscreened and screened predictor networks. Predictors retained by the screening procedure were those that exhibited a significant linear correlation with the raw Niño-3.4 index over the calibration period. To gauge significance, we performed Monte Carlo tests with an ensemble of 1000 first-order autoregressive [AR(1)] processes with identical lag-1 autocorrelation γ as the proxy of interest. The absolute value of the correlation between Niño-3.4 and each red-noise series is set aside; the 90th quantile of their distribution then defines an approximate 90% confidence bound against the null hypothesis that the absolute correlation is zero.

The effect of screening further reduces the number of proxy predictors: with a HadSST2i target, the number goes from 57 to 36 series (see Fig. 2 of Part II). Of those 33 time series, only 3 go back to A.D. 1000. The first is the PC1 of the Palmer Drought Severity Index (PDSI) reconstruction from the North American Drought Atlas (NADA; Cook 2008), including only grid points equatorward of 35°N.³ The second is derived from a network of drought-sensitive trees from Chile (Le Quesne et al. 2006). The third is the Lake Challa varve thickness record of Wolff et al. (2011).

b. Reconstruction methods

There is now a rich history of climate reconstruction techniques from proxy records; we refer to Jones et al.

(2009) for a recent and comprehensive review. While few records are available from the tropical Pacific (strictly defined), one can take advantage of the large-scale teleconnectivity of tropical climate to observe ENSO behavior from remote locations (Evans et al. 2002). Indeed, historical use of the term “teleconnection” itself is intimately tied to the discovery of the ENSO phenomenon (Walker and Bliss 1932; Bjerknes 1969; Wallace and Gutzler 1981; Horel and Wallace 1981) and its physical basis is well understood (e.g., Hoskins and Karoly 1981; Sardeshmukh and Hoskins 1988; Ting and Held 1990; Trenberth et al. 1998; Liu and Alexander 2007; Seager et al. 2003, 2005). The tropics (particularly in the circum-Pacific region) are dotted with ENSO-sensitive sites (Ropelewski and Halpert 1987) where paleoclimate archives can be successfully used to monitor past ENSO conditions (Evans et al. 2000, 2002). This was also the approach taken in early ENSO reconstructions (Quinn 1993; Whetton and Rutherford 1994), as well as more recent attempts (Gergis and Fowler 2009; Braganza et al. 2009), albeit in a more qualitative sense. Climate reconstruction techniques make use of the same physical relationships but allow climate anomalies to be quantitatively assessed from paleoclimate proxies, together with an estimate of uncertainty. They rest on the assumption that each proxy records some aspect of tropical Pacific SSTs in an approximately linear manner, and with a stationary covariance structure. These assumptions are testable and seem to hold to first order (Evans et al. 2000; Sterl et al. 2007; Furtado et al. 2009).

1) HYBRID REGEM RECONSTRUCTION

Following recent work (Mann et al. 2007b, 2008; M09), we use the RegEM technique (Schneider 2001) in a hybrid context, that is, reconstructing separately the high- and low-pass-filtered data (with a frequency split $f_s = 0.1 \text{ yr}^{-1}$). RegEM is a variant of the expectation maximization (EM) algorithm (Dempster et al. 1977; Little and Rubin 2002) designed for the imputation of missing values in spatiotemporal datasets typically encountered in climatology. It consists of regressing missing values over available ones using maximum-likelihood estimates of the mean and covariance matrix of the data. Its central assumptions are that (i) the data are missing at random, in the sense that the probability of a data point being missing is conditionally independent of the value of the measurement, and (ii) the data are normally distributed.

The RegEM algorithm has been widely applied to paleoclimate reconstructions (Mann and Rutherford 2002; Rutherford et al. 2003; Mann et al. 2005; Rutherford et al. 2005; Mann et al. 2007b, 2008; Riedwyl et al. 2008;

³ This series is closely related to the NADA PDSI PC1 series used by Li et al. (2011) to reconstruct ENSO variance over time, the only difference being the latitudinal restriction.

M09; W10). Smerdon and Kaplan (2007) note that LF variability is poorly preserved in RegEM when using Tikhonov regularization [also known as “ridge regression”; Tikhonov and Arsenin (1977)] but it was shown that regularization by TTLS (Fierro et al. 1997) is more robust in this regard (Mann et al. 2007a). A downside to TTLS is that, unlike ridge regression, a discrete truncation number—the number of modes retained in the singular-value decomposition of the augmented data matrix—must be a priori specified for the proxy network and time interval at hand. Truncation parameters for both high- and low-frequency networks were assigned for each 100-yr interval (A.D. 1000–99, A.D. 1100–99, . . . , A.D. 1800–50 based on the criteria of Mann et al. (2008)). The HF truncation level is chosen as the number of “significant” principal components of the proxy data matrix in each 100-yr interval (estimated as the number of singular values of the proxy data matrix that lie above a least squares fit to the trailing singular values in a log-linear scree plot). The LF truncation level is chosen so as to retain at least 33% of the variance in this spectral range.⁴ Hence the truncation choice is broadly adaptive and reflects the need for increased regularization back in time, when imposed by data sparsity (see Fig. 2b). Just as minor departures from Gaussianity do not invalidate normal statistics altogether, the fact that the missing-at-random assumption is not strictly satisfied in the paleoclimate context (Smerdon et al. 2008; Tingley et al. 2012) has been shown by the pseudoproxy studies cited above, and this study (section 5), to yield reasonable reconstructions nonetheless.

2) COMPOSITE-PLUS-SCALE RECONSTRUCTION

We also apply the relatively simple CPS methodology (Bradley and Jones 1993; Esper et al. 2002; Jones and Mann 2004; D’Arrigo et al. 2005; Wilson et al. 2006; D’Arrigo et al. 2009) to our proxy network, following W10, who underscored the value of CPS in reconstructing ENSO. They restored the mean and variance of each nest⁵ to those of the most replicated proxy nest, resulting in built-in homoskedasticity (maintenance of variance through time). This was done to avoid the loss of variance associated with the declining number of proxy predictors further back in time, but this artificial inflation of variance also increases reconstruction bias. Nevertheless CPS serves as a straightforward counterpart to RegEM, and will prove to provide comparable

estimates of decadal variability. To ensure meaningful averaging, we require that at least four proxy series be available at any given time. This is only the case after A.D. 1146, so our CPS reconstructions are zero beforehand.

4. Validation using frozen network analysis

A common way to assess the quality of statistical predictions is cross-validation (e.g., Hastie et al. 2008). In our context, it consists of estimating the instrumental target (here the historical Niño-3.4 index) using proxy data, when some of the instrumental data is withheld from the calibration period to compute out-of-sample prediction error (Wilks 1995). When the diminishing availability of data with time is taken into account, this validation exercise is called a frozen network analysis (FNA; see, e.g., Mann et al. 2007b). Following McShane and Wyner (2011), we selected contiguous, sliding holdout blocks as verification sets. The blocks are 45 years long and were slid in 2-yr steps, resulting in 48 blocks covering the A.D. 1857–1995 interval (the instrumental interval available in all three SST datasets). Calibration sets are defined as the complement of the verification set in the instrumental set; for example, an A.D. 1931–75 verification interval corresponds to a [A.D. 1857–1930] \cup [A.D. 1976–95] calibration interval. Compared to traditional calibration/verification exercises (which consider at most two disjoint intervals), this allows the estimation of a distribution of verifications scores for each network. We note that only about three truly independent 45-yr blocks can cover a 145-yr interval, so the effective sample size is lower than 48. We consider nine time segments to define the networks: A.D. 1000–99, A.D. 1100–99, . . . , A.D. 1700–99, and A.D. 1800–50.

To ensure that no information from the verification sets enters the prediction of Niño-3.4, the proxy screening process is repeated separately for each block, using calibration data only. Following standard practice, we report three statistics as an objective measure of quality 1) the reduction of error (RE), 2) the “coefficient of efficiency” (CE; Briffa et al. 1988; Cook et al. 1994),⁶ and 3) the coefficient of determination R^2 (e.g., Wilks 1995) (the square of Pearson’s product moment correlation). For observed and estimated time series z and \hat{z} ,

⁴ Mann et al. (2008) actually used a 50% cutoff, but our tests and those of M09 proved 33% to yield more robust reconstructions.

⁵ Defined as an interval over which proxy availability stays constant.

⁶ As noted by Bürger (2007), this statistic is not in fact the “coefficient of efficiency” proposed by Nash and Sutcliffe (1970), but has nevertheless become known under this name in dendrochronology and paleoclimatology.

TABLE 1. Results of the FNA: median scores. The reconstructions were performed on all three instrumental targets using screened networks. Header indicates a nest's starting year. Critical values ("Crit") reported here are the 95% quantiles of distributions obtained with AR(2) predictors fit to each instrumental target.

Case	Metric	1800	1700	1600	1500	1400	1300	1200	1100	1000	Crit
High frequency											
ERSSTv3, TTLS	RE	+0.73	+0.69	+0.59	+0.52	+0.58	+0.61	+0.67	+0.64	+0.48	−0.22
	CE	+0.73	+0.69	+0.59	+0.52	+0.58	+0.61	+0.67	+0.64	+0.48	−0.22
	R^2	+0.73	+0.69	+0.61	+0.56	+0.62	+0.62	+0.69	+0.67	+0.53	+0.15
HadSST2i, TTLS	RE	+0.73	+0.69	+0.60	+0.52	+0.59	+0.67	+0.65	+0.65	+0.47	−0.24
	CE	+0.73	+0.69	+0.60	+0.52	+0.59	+0.67	+0.65	+0.65	+0.47	−0.25
	R^2	+0.73	+0.70	+0.61	+0.54	+0.62	+0.67	+0.69	+0.67	+0.54	+0.15
Kaplan, TTLS	RE	+0.73	+0.73	+0.65	+0.54	+0.61	+0.66	+0.68	+0.66	+0.44	−0.22
	CE	+0.73	+0.73	+0.65	+0.54	+0.61	+0.66	+0.68	+0.65	+0.44	−0.22
	R^2	+0.75	+0.74	+0.69	+0.57	+0.62	+0.67	+0.73	+0.71	+0.53	+0.15
Low frequency											
ERSSTv3, TTLS	RE	−0.24	−0.09	+0.17	−0.46	−0.41	−0.33	+0.16	−0.13	−0.57	+0.20
	CE	−0.31	−0.22	−0.14	−1.36	−1.25	−1.05	+0.03	−0.33	−1.20	+0.02
	R^2	+0.13	+0.20	+0.36	+0.23	+0.18	+0.29	+0.39	+0.29	+0.24	+0.54
HadSST2i, TTLS	RE	−0.08	−0.01	+0.42	−0.27	−0.24	−0.08	+0.34	+0.07	−0.32	+0.16
	CE	−0.15	−0.14	+0.29	−0.46	−0.41	−0.17	+0.27	+0.02	−0.44	+0.06
	R^2	+0.19	+0.32	+0.48	+0.32	+0.22	+0.35	+0.49	+0.43	+0.34	+0.52
Kaplan, TTLS	RE	+0.08	+0.12	+0.24	+0.26	+0.16	+0.27	+0.18	+0.23	+0.04	+0.23
	CE	−0.15	−0.19	+0.08	+0.14	+0.04	+0.12	+0.11	+0.12	−0.22	+0.11
	R^2	+0.35	+0.37	+0.37	+0.34	+0.30	+0.37	+0.48	+0.49	+0.36	+0.53
Total											
ERSSTv3, TTLS	RE	+0.66	+0.60	+0.56	+0.41	+0.47	+0.54	+0.63	+0.57	+0.34	−0.28
	CE	+0.65	+0.60	+0.54	+0.39	+0.46	+0.50	+0.62	+0.54	+0.31	−0.31
	R^2	+0.70	+0.67	+0.58	+0.51	+0.55	+0.62	+0.66	+0.64	+0.49	+0.13
HadSST2i, TTLS	RE	+0.69	+0.66	+0.60	+0.46	+0.52	+0.58	+0.65	+0.61	+0.38	−0.31
	CE	+0.68	+0.65	+0.60	+0.46	+0.51	+0.57	+0.64	+0.60	+0.38	−0.32
	R^2	+0.73	+0.70	+0.61	+0.49	+0.57	+0.64	+0.68	+0.65	+0.53	+0.14
Kaplan, TTLS	RE	+0.69	+0.67	+0.62	+0.51	+0.54	+0.60	+0.64	+0.61	+0.43	−0.31
	CE	+0.69	+0.67	+0.61	+0.50	+0.53	+0.59	+0.63	+0.61	+0.41	−0.34
	R^2	+0.73	+0.73	+0.66	+0.53	+0.56	+0.62	+0.69	+0.67	+0.50	+0.13

$$\text{RE} = 1 - \frac{\sum_{i=1}^N (z - \hat{z})^2}{\sum_{i=1}^N (z - \mu_c)^2} \quad \text{and} \quad (1)$$

$$\text{CE} = 1 - \frac{\sum_{i=1}^N (z - \hat{z})^2}{\sum_{i=1}^N (z - \mu_v)^2}, \quad (2)$$

where μ_c and μ_v are the means of z over the calibration and verification intervals, respectively. A perfect reconstruction will exhibit unit scores, but there is debate as to which scores delineate a "skillful" reconstruction. Ammann and Wahl (2007) point out that R^2 is often an inadequate metric of LF skill, being overly prone to false negatives and false positives, so its results should be viewed with caution. The most stringent of the three metrics is CE, which by definition is always less than RE. B rger (2007) showed that CE, which rewards

reconstructions tracking shifts in the mean with appropriate amplitude, derives from a skill score (Wilks 1995). RE does not, but has a more natural interpretation as the amount of variance captured by a reconstruction, with the caveat that stationary stochastic processes can produce positive RE scores by chance alone. To gauge the influence of persistence on validation scores, we perform 1000 Monte Carlo simulations with random time series fitted to the instrumental target (at both high and low frequencies). The chosen model is a second-order autoregressive [AR(2)] method, chosen using the "ARfit" algorithm (Schneider and Neumaier 2001). From these distributions we extract the 95% quantile, yielding the critical values RE_{crit} , CE_{crit} , and R^2_{crit} (Table 1).

a. RegEM validation

Distributions of RE scores for the frozen network analysis are shown in Fig. 3 for a HadSST2i target. Total scores are seen to remain relatively constant over time, in the vicinity of 0.5. This is primarily explained by the

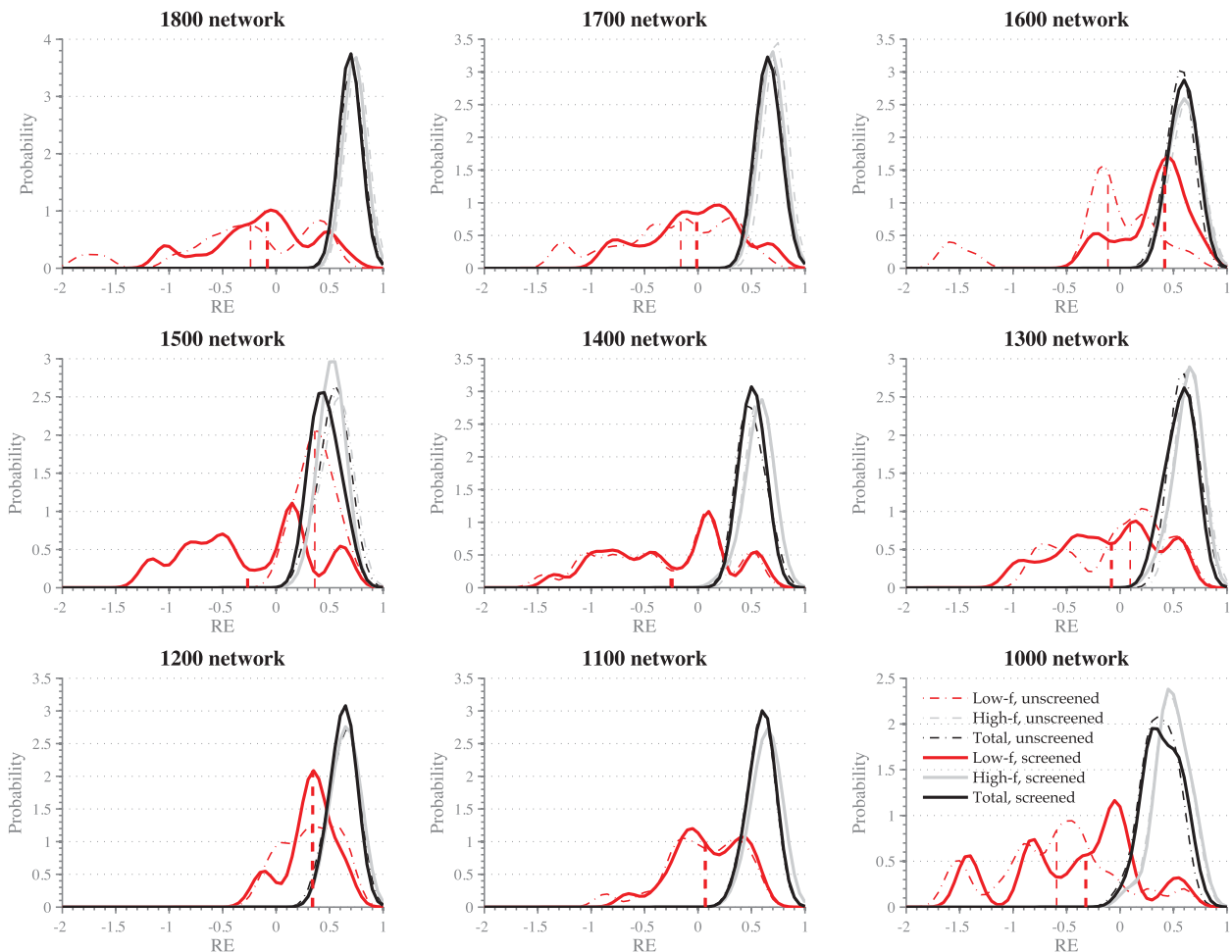


FIG. 3. The FNA of RE scores of RegEM HadSST2i reconstructions, for century-by-century networks and by spectral band (LF, HF, and total). Plots represent distributions obtained by kernel density smoothing of the raw histogram, with a bandwidth of 0.1. Solid lines represent scores obtained with screened proxy networks, while dash-dotted lines correspond to unscreened networks. Vertical dashed lines depict the median of the LF distributions.

large HF scores, while LF scores fluctuate widely. The fact that median RE scores do not increase monotonically as proxy networks grow denser may be due to several factors: colinearity (redundancy) between proxies, overfitting to noise, or an overly simplistic statistical model—we cannot distinguish among these possibilities. Solid lines represent scores obtained with screened proxy networks, and generally show more skill in the post-A.D. 1600 period, which means that these problems are partially mitigated by the screening procedure. Still, screened-network LF RE scores only breach zero for three networks: A.D. 1100, A.D. 1200, and A.D. 1600. A potential problem in the assessment of LF skill by FNA is the shorter calibration period, reduced from 145 to 100 yr (ERSSTv3 and HadSST2i) and from 139 to 94 yr (Kaplan) compared to a real reconstruction. This question will be explored using pseudoproxy experiments in section 5c, where it will be argued that LF scores are systematically

underestimated by FNA compared to what would be obtained from longer calibration and verification periods, and the uncertainties are thereby increased.

A more comprehensive picture is offered in Fig. 4, which assesses screened reconstructions for all three targets. Since R^2 often fails to discriminate between good and poor reconstructions, only the median RE and CE scores are reported here for conciseness, although all three statistics are reported in Table 1. Sensitivity tests (not shown) demonstrate that the results are not sensitive to small (5–10 yr) variations in the length of the verification period or the overlap between blocks, as long as a few decades are available for verification.

Several results are noteworthy:

- Critical values of RE and CE are usually negative, except at low frequencies, where they reach up 0.23

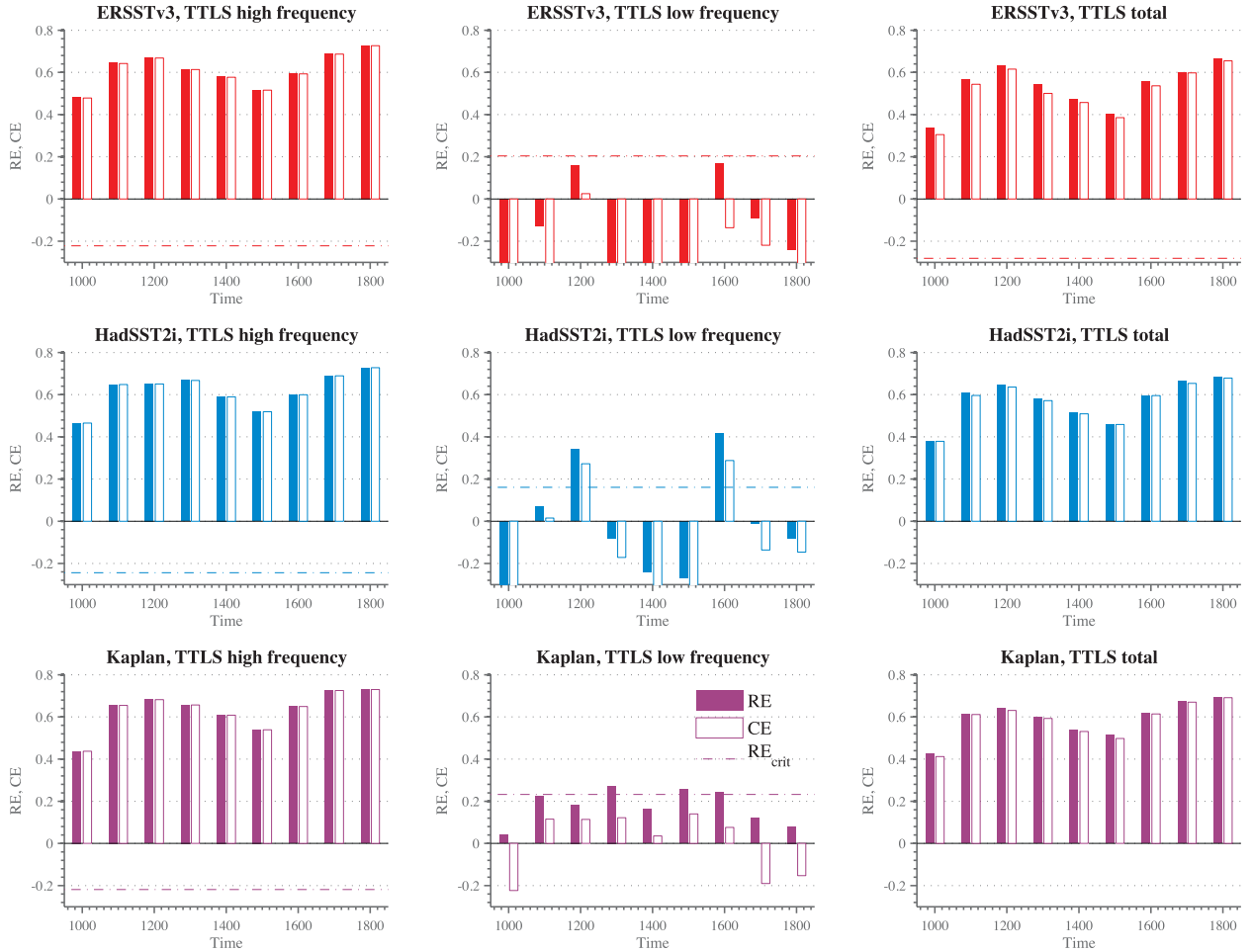


FIG. 4. The FNA of median verification scores for all three targets. Solid bars represent the RE statistic, and open bars represent the CE statistic. The RE_{crit} is depicted by the dashed lines, driven by the persistence of each Niño-3.4 index. Because CE_{crit} is always negative, it is not shown. Any score above the critical value indicates a skillful reconstruction by this metric.

(Kaplan). Critical R^2 scores are on the order of 0.5 for LF, and 0.13 for total scores.

- Consistent with previous work, verification scores do not necessarily improve as denser networks of predictors are used, confirming the impression gleaned from Fig. 3. The RE or CE scores do not increase monotonically over time with any target.
- The HF verification scores considerably exceed the critical values for all metrics for all networks. This generally good performance is due to the fact that Niño-3.4 displays most of its variance in the interannual band, which is also where the signal-to-noise ratio is highest in most proxy measurements. Accordingly, we find that total scores are always above critical values even when LF scores are not, which may overstate our true confidence in the reconstructions' quality given the presence of age uncertainties in the preinstrumental era.

As a consistency check, we analyzed regression residuals for randomly selected verification periods. Residuals were tested for Gaussianity via the Spiegelhalter test (Spiegelhalter 1983) and for autocorrelation via the Durbin–Watson test (Durbin and Watson 1951; Kanzler 1998) (not shown). Skillful reconstructions generally exhibit test p values exceeding 0.05, indicating that residuals are statistically indistinguishable from uncorrelated, Gaussian residuals at the 95% level. Conversely, reconstructions that do not pass the CE or RE thresholds consistently exhibit non-Gaussian and autocorrelated residuals, confirming the poor quality of such predictions.

b. CPS validation

In Fig. 5 we plot the RE score for the CPS reconstructions. By design, CPS rescales the composite in each nest (which changes every time a proxy drops in

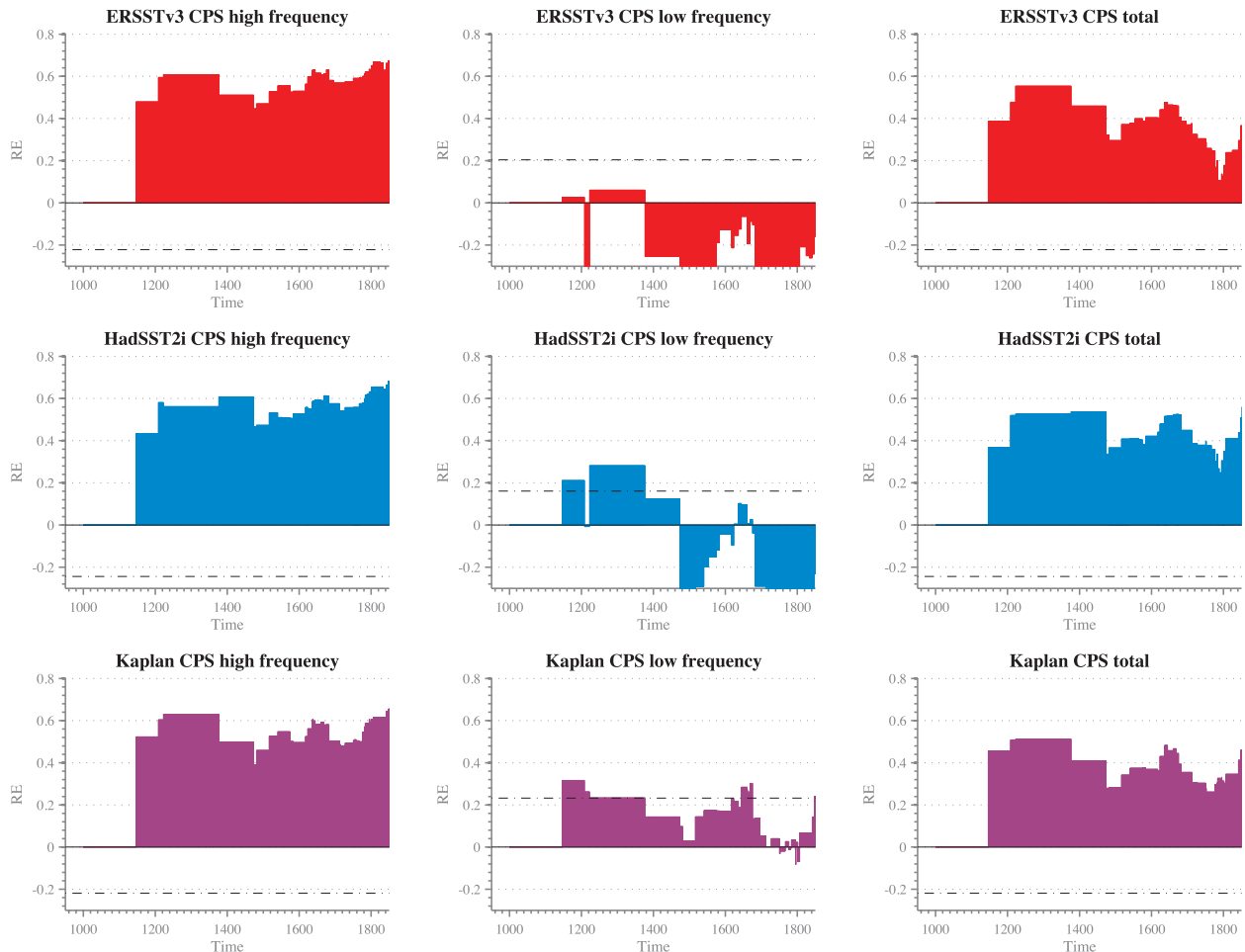


FIG. 5. As in Fig. 4, but using FNA of Niño-3.4 CPS reconstructions for RE only.

or out of the network), so an RE value is computed for each such nest (sometimes as short as a few decades). The reconstructions display consistently positive HF RE scores (about 0.6, like RegEM reconstructions) for all times after A.D. 1150. In contrast, LF RE scores are mostly insignificant, indicative of skill-less reconstructions—a result found to stem from poor estimates of the mean in each nest. That CPS should fail to reconstruct centennial variability is not surprising given that it centers composites at every nest, thereby erasing the previous nest's memory with regard to the mean. Therefore, one can only expect CPS to perform well on time scales shorter than the average nest size. For centennial variability to be skillfully reconstructed with CPS, one would need to employ proxy series available continuously through several centuries, not on a network characterized by a progressive dropout of proxy series back in time, such as ours (Fig. 2). Furthermore, CPS implicitly assumes that each proxy has an equal say in

the final reconstruction, and that the central limit theorem will take care of reducing errors. Since every proxy expresses Niño-3.4 variability in different ways, the very structure of the regression framework enables it to capture such differences, unlike CPS. These limitations lead us to favor RegEM for LF reconstructions, though CPS reconstructions will provide a useful reference point in Part II.

c. Synthesis

Overall, we derive four major conclusions from this exercise:

- 1) In the absence of dating errors, our network enables the skillful reconstruction of Niño-3.4 interannual variance with RegEM and CPS, typically around 60% thereof. In accordance with previous studies, we find LF variability relatively more difficult to reconstruct because most of the signal is in the interannual (ENSO) band.

- 2) The RegEM solutions capture LF variability more skillfully than CPS in all cases, but LF scores vary strongly with the target SST dataset.
- 3) For all three targets, screened networks generally improve prediction skill.

Differences in median total RE scores are overall fairly minor (Fig. 4, rightmost column), and one may ask whether they are statistically significant given the width of the distributions (illustrated in Fig. 3). A caveat of frozen network approaches is that the length of the calibration period is reduced from ~ 150 to ~ 100 yr, which affects both the screening process (tending to select a slightly different list of proxies than when screening over the entire historical interval) and the ability to capture centennial-scale variability. These considerations may result in overly conservative conclusions in regards to LF skill. Conversely, neglecting the dating errors likely leads to an overly optimistic assessment of HF skill. We now turn to pseudoproxy experiments for a complementary appraisal of our methodology.

5. Validation using pseudoproxy experiments

a. Experimental design

Although the previous analysis constitutes an important crosscheck of the methodology, it only evaluates its performance vis-à-vis the decreasing number of predictors back in time (i.e., the network depopulation). Pseudoproxy experiments (PPEs; Smerdon 2012) offer another useful benchmark: the numerical laboratory provided by long integrations of coupled general circulation models (CGCMs), wherein the effects of stochasticity, varying proxy quality, and frequency-specific calibrations can be investigated (Mann and Rutherford 2002; von Storch et al. 2004; Mann et al. 2005, 2007b; Hegerl et al. 2007; Küttel et al. 2007; Riedwyl et al. 2008; Smerdon et al. 2010b, 2011).⁷

Here we use the pseudoproxy framework to test the robustness of the hybrid RegEM reconstruction of Niño-3.4, using the output of two 1156-yr-long GCM simulations with the NCAR Community Climate System Model, version 4.0 (CCSM4.0), CGCM:⁸

- 1) LM: A simulation of the last millennium forced with estimates of volcanic and solar forcing, covering the A.D. 850–2005 period.
- 2) PI: An unforced simulation of the last millennium, using fixed A.D. 1850 top-of-the-atmosphere conditions.

These components follow phase three of the Paleoclimate Modelling Intercomparison Project (PMIP3) protocols for last millennium experiments described in Schmidt et al. (2012). Their implementation in CCSM4.0 is documented by Landrum et al. (2013). The model's simulation of tropical Pacific climate is described extensively in Deser et al. (2012). It is characterized by a realistic seasonal evolution of atmospheric and oceanic fields, realistic asymmetry between El Niño and La Niña events, and a realistic pattern of teleconnections. Interannual ENSO amplitude, however, is overestimated by approximately 30%, while the decadal component is weaker than observed and displays subdued extratropical linkages. The LF variability in the PI control simulation must be intrinsically generated by the coupled system, since it is not subjected to changing external forcings such as greenhouse gas concentration, volcanic aerosols, or solar irradiance variations. This permits an assessment of Niño-3.4 reconstruction skill in the face of intrinsically generated global variability only.

Simulated Niño-3.4 spectra are plotted in Fig. 6, alongside instrumental Niño-3.4 spectra from the three historical SST analyses under consideration. In addition to the points mentioned above, it can be seen that the three instrumental spectra indeed fall in between the LM and PI end members at centennial scales, and that the magnitude of their centennial variability scales like the linear trends of Fig. 1. The greater multidecadal variability in the instrumental records could be due in part to anthropogenic aerosol forcings, which are omitted from the PI control simulation and enter into the LM simulation only at the end of the time series. The ENSO-band variability is quite realistic in both simulations, but overly energetic at the quadrennial scale, particularly in LM. This greater ratio of HF (interannual) to LF variability in the simulations must be kept in mind to interpret our results.

In seeking to construct meaningful pseudoproxy experiments to test our methodology, a few points deserve consideration:

- Contrary to the usual assumption used in PPEs, many proxies in our network were selected on the basis of their correlation to Niño-3.4 SST, not local SST. This is particularly true of coral $\delta^{18}\text{O}$ records, which often record stronger hydrological than thermal signals (Gagan et al. 2000) but can nevertheless faithfully reconstruct large-scale SST patterns by virtue of the

⁷ We note that an error in model gridbox locations impacted conclusions of Mann et al. (2005, 2007b) bearing specifically on Niño-3 reconstructions (Smerdon et al. 2010a), but it left the overall conclusions of the study unchanged (Rutherford et al. 2013). None of those results are relevant to the present study and so should not be used as a point of comparison.

⁸ To be consistent with our reconstruction choices, only the A.D. 1000–1995 period DJF averages will be analyzed here.

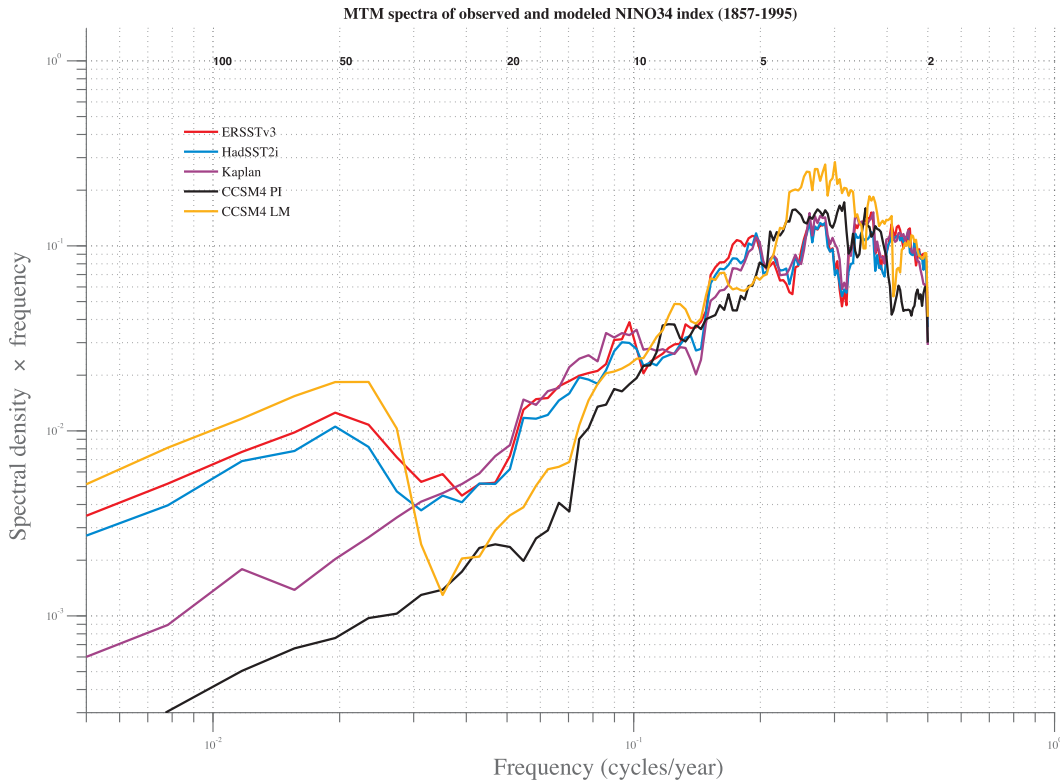


FIG. 6. Multitaper spectrum of Niño-3.4 index in the CCSM4.0 LM and PI simulations, together with spectra of the corresponding index derived from historical SST analyses (ERSSTv3, HadSST2i, and Kaplan) over the period A.D. 1857–1995. The spectra are variance preserving, in the sense that, upon transforming the ordinate to a linear scale, the area under each curve would sum to the total variance. They were obtained via the multitaper method of Thomson (1982) with a half-bandwidth of 4. Numbers across the top refer to the period in years.

strong coupling between tropical SSTs and rainfall (Evans et al. 1998). Our network must thus reflect this diversity of physical controls.

- In the absence of forward models calibrated to each proxy, a simplifying assumption must be made to construct pseudoproxies P at each location x and time t . We follow previous studies (e.g., Mann et al. 2007b; Christiansen et al. 2009; Smerdon et al. 2010b, 2011) in using a white-noise model:

$$P(x, t) = V(x, t) + \xi(x, t)/\text{SNR}, \quad (3)$$

where $V(x, t)$ is the controlling physical variable, $\xi(x, t)$ is a realization of a Gaussian white noise with zero mean and unit variance, and SNR is the signal-to-noise ratio (SNR = ∞ , 1.0, 0.5, and 0.25).

- The $V(x, t)$ was assigned as follows: all corals are assumed to record local temperature; ice cores and speleothems are assumed to record (power transformed) precipitation; tree-ring records are assumed to record drought conditions as described by the PDSI; Lake Challa varve thickness (Wolff et al. 2011) is controlled by zonal wind, while Cariaco basin

Mg/Ca (Black et al. 2007) is controlled by local temperature.

- By consistency with our methodology, we screen each pseudoproxy for a significant correlation to the target Niño-3.4.
- Given the rapid decrease in proxy availability over time (Fig. 2), and the marked temporal dependence of verification statistics (section 4), it is clear that any realistic pseudoproxy network should exhibit a similar pattern of temporal availability. To do this, we rank the screened pseudoproxies by their correlations to the simulated Niño-3.4. Similarly, we rank the screened real proxies by their correlations with HadSST2i Niño-3.4 SST over the historical epoch. Then we pair up the pseudo- and real proxies by rank, and assign the temporal mask from the real proxy to its rank-partner pseudoproxy. We generate a reconstruction using the masked pseudoproxies, as we do for real proxies. This process is repeated 200 times for each noise level.

These features make our PPEs more realistic than previous studies in some respects and less so in others.

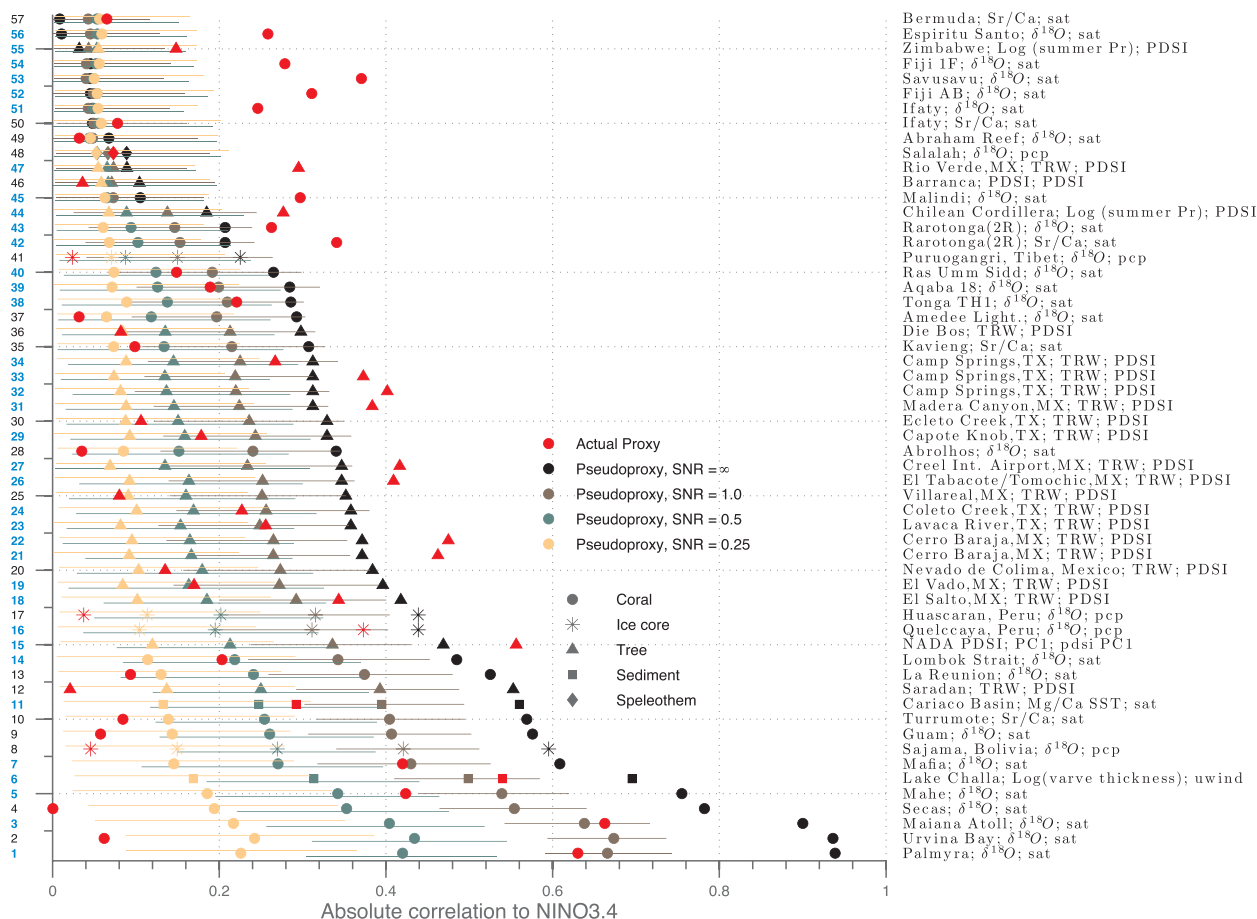


FIG. 7. Observed and simulated correlations with Niño-3.4. Symbols mark the median of the ensemble for pseudoproxies, while horizontal bars depict 95% confidence intervals about this median. Red symbols mark the absolute value of observed correlations (Table A1), and the shape of each symbol indicates proxy type. Records are ranked in decreasing order of $|\rho_\infty|$, the correlation of pseudoproxies to the CCSM LM Niño-3.4 index in the SNR = ∞ case. Numbers in cyan refer to those proxy records that were retained by the screening process (boldface numbers in Table A1). The text column on the rhs lists site; measurement type; and controlling variable. The TRW stands for tree-ring width. Controlling variables are surface air temperature (sat), PDSI, (power transformed) precipitation (pcp), and zonal wind speed (uwind). Note that pseudoproxy correlations do not correspond in detail to real-world correlations, although the case SNR = 1.0 is closest overall to capturing the evolution of absolute correlations through time (Fig. 9, described below).

On the one hand, the design takes advantage of the fact that real-world proxies record a range of physical variables (not just temperature) related to Niño-3.4 SST via teleconnections, which are allowed to evolve with the model's climate state. It is also the first (to our knowledge) to factor in real-world proxy availability. On the other hand, our setup radically simplifies the physical controls on each proxy, neglecting the multivariate nature of most proxy systems (with attendant constructive or competing effects between variables), especially those involving oxygen isotopes. While this design is far from ideal, it is a step toward a more comprehensive treatment, which must ultimately involve high-quality GCM simulations coupled to

validated forward models of proxy formation. At present, biased GCMs and incomplete (or non-existent) forward proxy models limit our ability to assess paleoreconstruction skill. However, our study emphasizes important questions that can help guide future work.

The synthetic and actual proxies are compared in Fig. 7, where it is evident that the pseudoproxy design tends to overestimate true correlations for some proxies and underestimate them for others. Note also that for sufficiently low correlations, the addition of noise to the local climate variable sometimes acts to increase absolute correlations to Niño-3.4 above their noise-free reference level. Hence, pseudoproxies constructed with

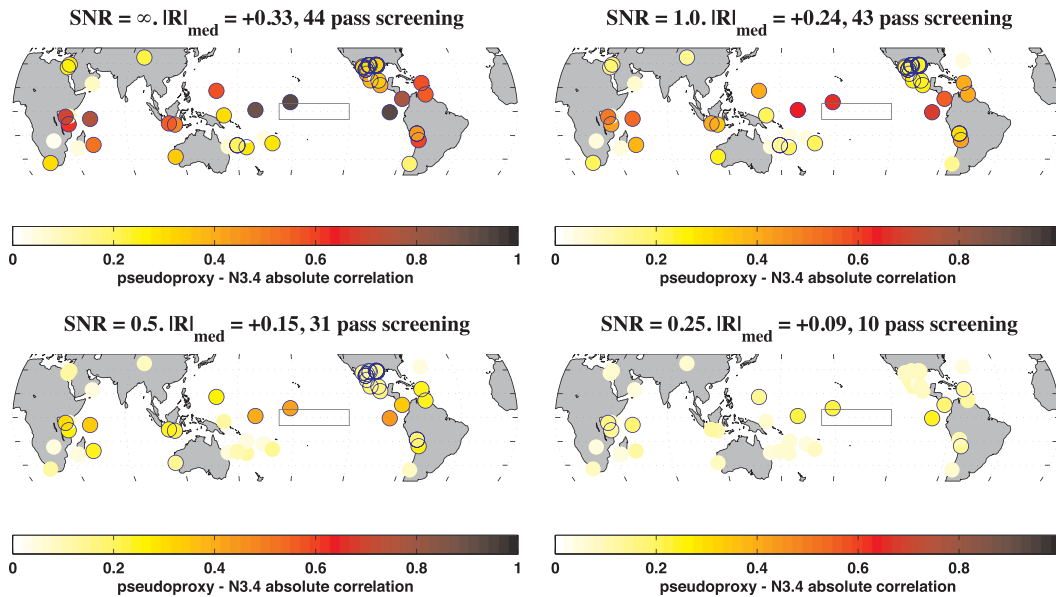


FIG. 8. Correlations of each pseudoproxy with Niño-3.4 for each SNR case (∞ , 1.0, 0.5, and 0.25) in the forced (LM) simulation. The color scale refers to the median absolute correlation. Circles outlined in black denote sites that exhibit significant correlations in at least 100 of the 200 noise realizations.

SNR = 1.0 may exhibit slightly higher correlations than those constructed with SNR = ∞ for proxy indices above 20 or so, although their median value is much lower than for SNR = ∞ . In the extreme case where noise-free correlations are close to nil (index above 45 or so), the majority of noise realizations for all other SNRs exhibit higher absolute correlations than for SNR = ∞ ; this is, however, not sufficient to make them significant, so they are always screened out of the pseudoproxy pool and so do not contribute to the reconstructions.

A spatial view of such correlations is presented in Fig. 8. As expected, the higher the SNR, the more proxies prove significant in each noise realization. Given the discrepancies noted above, however, one may ask which SNR case, if any, constitutes the closest analog to the observed Niño-3.4 signal strength. Overall, the histogram of observed absolute correlations to Niño-3.4 (not shown) is most comparable to the SNR = 1.0 case. This notion is confirmed by looking at the effective SNR over time—that is, the collective ability of the pseudoproxy network to capture a relationship to Niño-3.4 over time in a manner that mimics reality. To quantify this, we define the partial SNR as

$$\text{SNR}_{\text{Niño-3.4}}^{(i)} = \frac{r_i}{\sqrt{1 - r_i^2}}, \quad (4)$$

where r_i is the absolute correlation of the i th proxy (or pseudoproxy) to Niño-3.4. This expression—only valid

for standardized variables—is a direct consequence of the definition of the pseudoproxies and the linearity of the correlation operator. The effective SNR is then the sum of all partial SNRs for all proxies available at a given point in time. It is plotted in Fig. 9, which shows the network SNR = 1.0 to be most comparable to the observed network.

b. Results: High- and low-frequency performance

We compute the RE and CE statistics for the raw and low-pass-filtered time series over the A.D. 1001–1850 verification interval (with respect to an A.D. 1851–1995 calibration), and display the results in Fig. 10 for the LM simulation and in Fig. 11 for the PI simulation. As in the previous section, we tested our results against distributions drawn from an ensemble of noninformative predictors, taken as 1000 realizations of an AR(2) process fitted to the target Niño-3.4. Hence, a good reconstruction should exhibit both RE > 0 and beat the noninformative prediction (in the sense that its sampling distribution of RE and CE scores should be to the right of the noninformative prediction, with little to no overlap). To account for the effects of stochasticity, 200 random realizations of the noise matrix were generated, yielding 200 independent pseudoproxy tests. This large ensemble size is important in order to derive the most general conclusions (Christiansen et al. 2009; see also Rutherford et al. 2010).

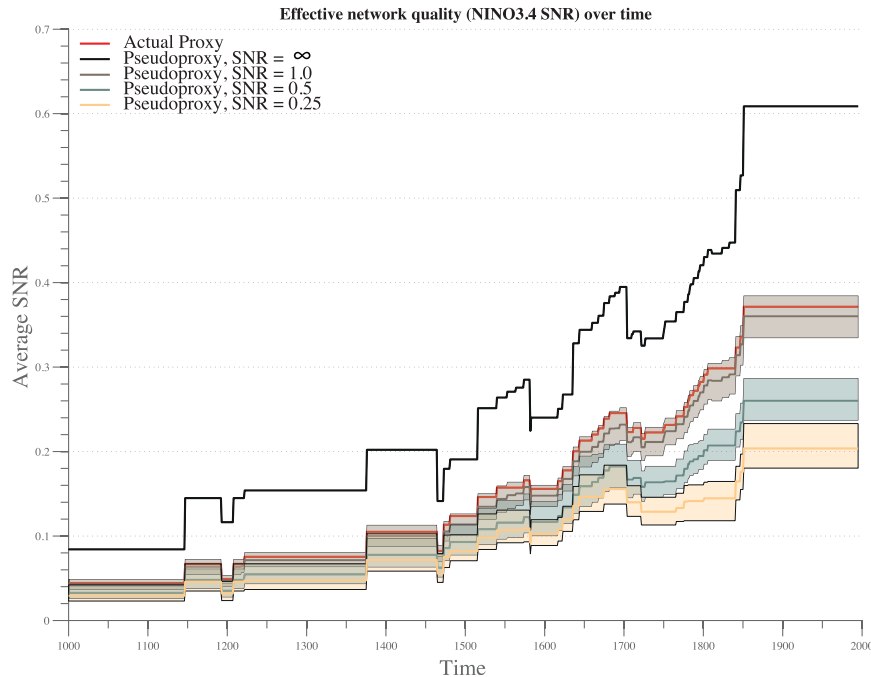


FIG. 9. Effective network quality expressed as the Niño-3.4 SNR [Eq. (4)] for the observed and four pseudoproxy networks (after infilling the data matrix over the instrumental interval to ensure homogeneity). For the observed SNR, absolute correlations were computed using the HadSST2i Niño-3.4 index.

In the LM case, the total and LF RE scores beat noninformative predictors at reconstructing Niño-3.4 in all cases but $\text{SNR} = 0.25$, and the $\text{SNR} = 1.0$ rivals or even exceeds the $\text{SNR} = \infty$ case (Fig. 10). That performance should increase despite a decrease in SNR deserves an explanation: it is a consequence of the fact, discussed above, that sampling variability may at times increase a noisy pseudoproxy's correlation to Niño-3.4 above its noise-free value, hence marginally improving the effective network quality. The case most comparable to the real network ($\text{SNR} = 1.0$) displays relatively high LF RE scores (median = 0.75), but important variance losses are apparent, as is a tendency for a warm bias confirmed by the negative LF CE scores (median = -0.3 ; not shown). This bias is due to the presence of a strong trend over the calibration interval, which is the warmest period of the whole interval.

In the PI case, reconstruction skill is comparable for the raw Niño-3.4 (Fig. 11b) but drastically lower at LF (Fig. 11c). This is not surprising, as LF variability is subdued in the PI simulation, and hence is not detectable by a sparse pseudoproxy network. That total RE scores should be similar to the LM case underscores the dominance of interannual variability in the CCSM4.0, with or without forcing. Given that the magnitude of LF variability in all three instrumental products (Fig. 6) is bracketed by those two end members, we can expect

the LM case to provide an upper bound, and the PI case a lower bound, on LF skill.

To visualize performance in frequency space, Fig. 12 shows the multitaper method (MTM) spectra of the target and reconstructed Niño-3.4 indices for $\text{SNR} = 1.0$. In the LM case, a large overlap between the 95% confidence regions indicates that reconstructions accurately depict the target Niño-3.4 spectrum out to periods of approximately 10 yr. For all lower frequencies, the reconstructions systematically underestimate variability, as expected from the “regression dilution” phenomenon (Frost and Thompson 2000). This effect is likely to be at play in real-world reconstructions as well. In the PI case, the lack of substantial multidecadal to centennial variability is accurately captured by the reconstructions, albeit generally with incorrect phase, as attested by the poor reconstruction statistics (Fig. 11). The reconstructions also correctly capture the spectral signature of variability at periods shorter than 10 yr, but slightly underestimate it in the 10–20-yr range.

Extending these results to the real world raises a few issues. First, the pseudoproxy model used here is a simplistic representation of how climate information is incorporated into proxy data. Second, there is no way of knowing which of the two GCM end members is closer to the true tropical Pacific climate. The existence

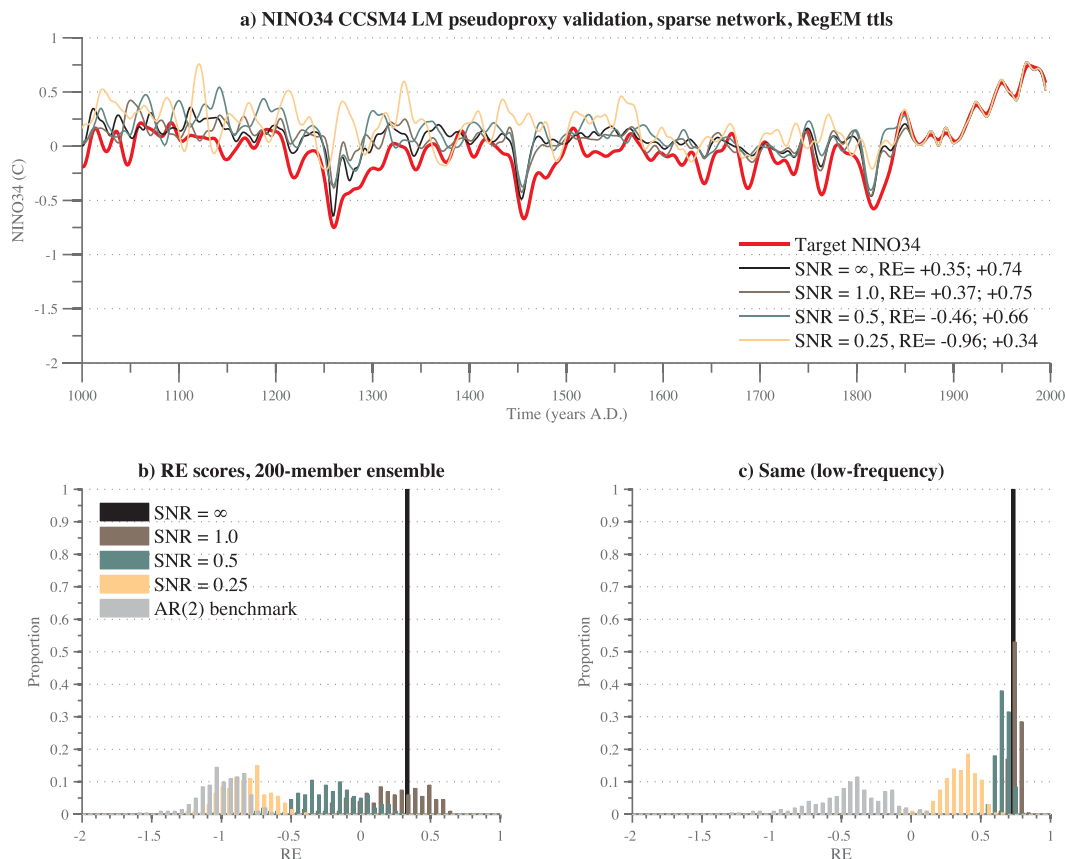


FIG. 10. Pseudoproxy validation of RegEM TTLS with CCSM4.0 LM target. (a) The 20-yr low-pass-filtered time series of target Niño-3.4 (red) and reconstructions based on the four proxy networks (SNR = ∞ , 1.0, 0.5, and 0.25) corresponding to the median of the LF RE distribution for each SNR. Numbers in the legend refer to total and LF RE scores for the plotted reconstruction, respectively. (b) Distribution of total RE scores. (c) As in (b), but for LF RE scores.

of important LF variability in many sedimentary records of ENSO (cf. section 1) tends to suggest that real variability was—regardless of its cause—closer in character to the LM-forced simulation than the PI simulation, implying that for realistic SNR values, RegEM TTLS should yield skillful reconstructions for both high and low frequencies using the available proxy network. Finally, dating uncertainties in real proxy records are expected to deteriorate HF (interannual) skill relative to the pseudoproxy case, but the effect is expected to be negligible at longer periods due to our stringent dating control (5 yr or finer). So while our reconstructions will likely underestimate the very long-term (millennial) trends, these results suggest that they should faithfully capture decadal to centennial oscillations, albeit with inevitable variance losses (i.e., for RE = 0.5, the amplitude will be damped by a factor of $\sqrt{2}$, but the phase will be correctly estimated). This claim of a faithful reconstruction of LF signals, however, is at odds with the frozen

network analysis of section 4, a problem to which we now turn.

c. Results: Accuracy of frozen network analyses

The PPE framework permits an evaluation of how FNA captures the true distribution of verification scores in an internally consistent manner. Here we wish to quantify the ability of calibration/verification exercises to quantify reconstruction skill when the answer is independently known and many noise realizations are available. To do this, we repeat the analysis of section 4 on the LM and PI Niño-3.4 targets, using the same protocol as above. We assume that only 145 years of “instrumental” data are available, partition them between 100-yr-long calibration set and 45-yr-long verification sets as before, and perform FNA on those sets. We then compare them to the true distributions of RE scores obtained with a 145-yr-long calibration and 100-yr-long verification intervals (A.D. 1000–99, A.D. 1100–99, ..., A.D. 1800–50). Because the two types of distributions are

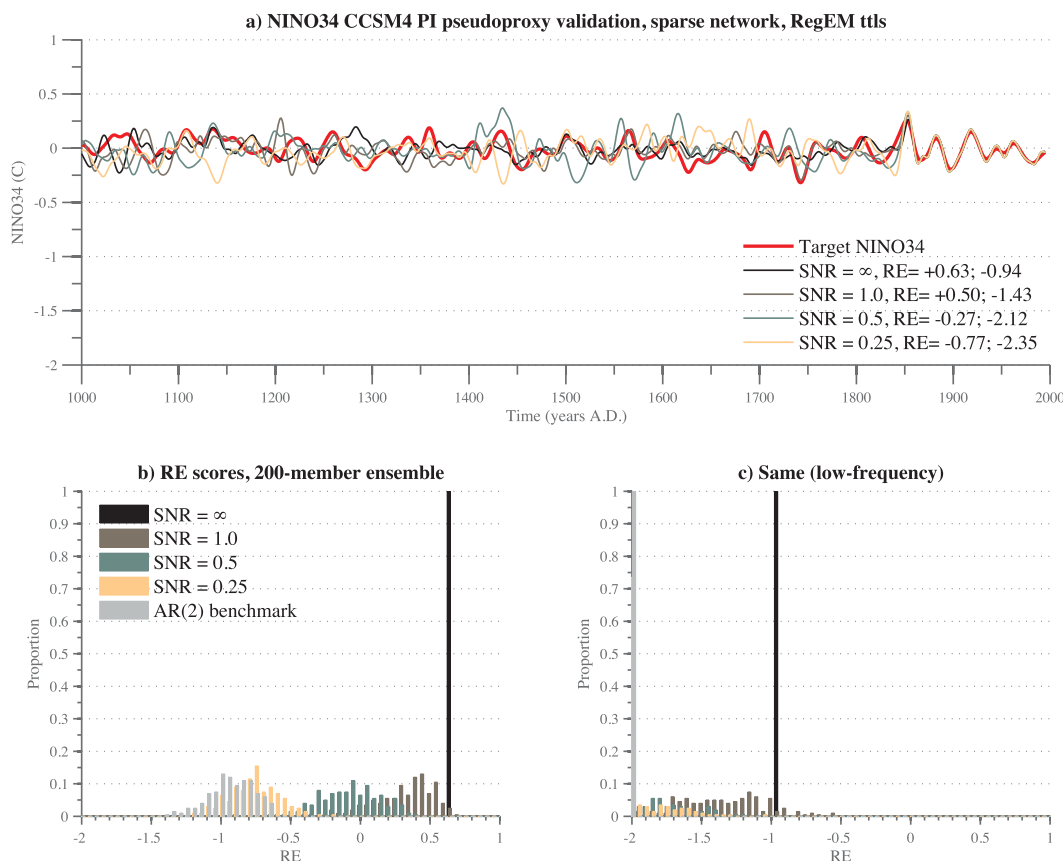


FIG. 11. As in Fig. 10, but with the CCSM4.0 PI target.

evaluated from samples of different size, we apply a kernel density smoothing with a bandwidth of 0.1 to remove granularity arising from sampling variability. For brevity, we only describe the SNR = 1.0 case, but results are qualitatively similar for all noise levels.

The LM results are shown in Fig. 13 for RE scores (similar conclusions hold for CE scores). Pseudoproxy (PP) verification scores (dash-dotted lines) are seen to worsen gradually over time, as expected from network depopulation (Fig. 9). Yet the FNA distributions (solid lines) are consistently shifted left of the PP distributions, especially at low frequencies, whereas HF distributions overlap largely with their PP counterparts, sometimes slightly under- or overestimating them. The only exceptions are the A.D. 1000–99 and A.D. 1100–99 nests, where FNA overestimates total skill. However, since their median never breaches zero, one would judge such reconstructions unreliable regardless of the assessment method. Interestingly, median LF PP RE scores for the forced simulation are high (>0.7) for all networks after A.D. 1200, so the total scores presented in Fig. 10 are brought down by the two earliest nests.

In the PI case (not shown), the opposite happens at LF: FNA tends to overestimate LF RE distributions, but in no instance do RE scores rise above zero, indicative of skill-less reconstructions. Because of the lack of substantial LF variability in this simulation, distributions of HF and total variability largely coincide, consistent with previous findings, and in all cases those scores are significantly underestimated by FNA compared to PP.

These results suggest that FNA can significantly err in its estimation of verification skill, especially at low frequencies. The case of most interest is LM, where LF variability is most comparable to that observed, and for which LF skill is consistently underrepresented in FNA compared to the PP case (by about 0.3 units over most nests). Although the aforementioned caveats limit these results' direct application to our real-world FNA (section 4), they suggest that reconstruction skill is generally underestimated by this assessment—presumably because of the low number of degrees of freedom available for calibration and/or verification. Note, however, that the FNA samples are not independent (two neighboring verification periods share

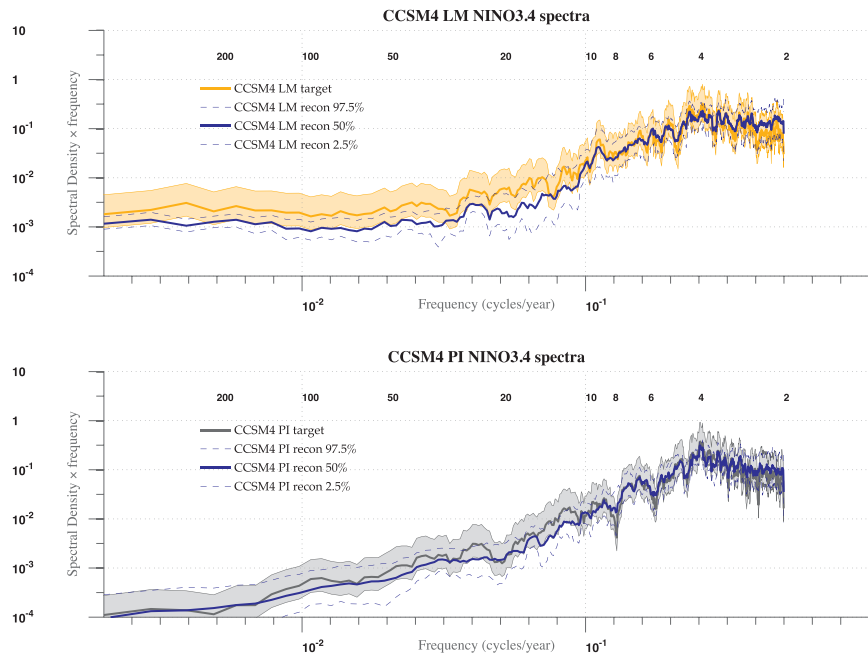


FIG. 12. Spectral characteristics of pseudoproxy reconstructions ($\text{SNR} = 1.0$). Variance-preserving MTM spectra are shown as in Fig. 6. Yellow and grey shaded regions represent error at a 95% confidence interval obtained via χ^2 estimates under ergodic, Gaussian assumptions. Blue dashed lines show the 95% confidence interval from the pseudoproxy ensemble of reconstructions, whose median is represented by the solid blue line. Numbers across the top refer to the period in years.

95% of data in common), so their size and spacing must reflect a compromise between sample size and the degree of overlap. We find that 30-yr holdout blocks advocated by McShane and Wyner (2011) are even less suited to the estimation of LF skill, and that increasing the spacing to 5 yr does not qualitatively affect our results. Overall, the chosen block size (45 yr) and spacing (2 yr) were found by trial and error to yield the least biased assessments of skill, although this bias remains strong. Finally, using an unscreened proxy network yields almost identical results, indicating that the short screening period used in FNA is not responsible for the lower LF skill.

6. Discussion

In this first part, we have set the stage for multiproxy reconstructions of tropical Pacific SST (Niño-3.4 index) over the past millennium. Using a network of accurately dated paleoclimate proxies back to A.D. 1000, we presented two largely independent methodologies (RegEM and CPS), which were validated by systematic cross-validation exercises and (for RegEM) pseudoproxy studies.

The frozen network experiments raised several key points: first, all methods used here are vulnerable to

noise in the data, so adding more proxy predictors does not necessarily reduce out-of-sample prediction error as measured by the RE, CE, or R^2 statistics. Efforts are underway to devise more robust and accurate statistical methods tailored to this problem. Second, we found LF (decadal and longer) variability inherently harder to reconstruct than HF (interannual) variability, consistent with previous studies (Gergis et al. 2006; Braganza et al. 2009; W10). Although we found that in all cases skillful LF reconstructions could be obtained back to A.D. 1200 with a HadSST2i or ERSSTv3 target (A.D. 1000 with a Kaplan target), FNA skill scores are far from uniform in time and are highly dependent on the instrumental target. This result is surprising and underscores the need for a better understanding of the divergence between SST data products at decadal to centennial time scales. Third, we found RegEM TTLS to be more skillful at reconstructing LF fluctuations than CPS, although HF skill was comparable between the two methods, with a slight advantage for RegEM. This leads us to favor the latter method in the millennium-long reconstructions of Part II. An important caveat of the frozen network framework is that it may compromise the quality of centennial-scale calibration by shortening the calibration and verification periods,

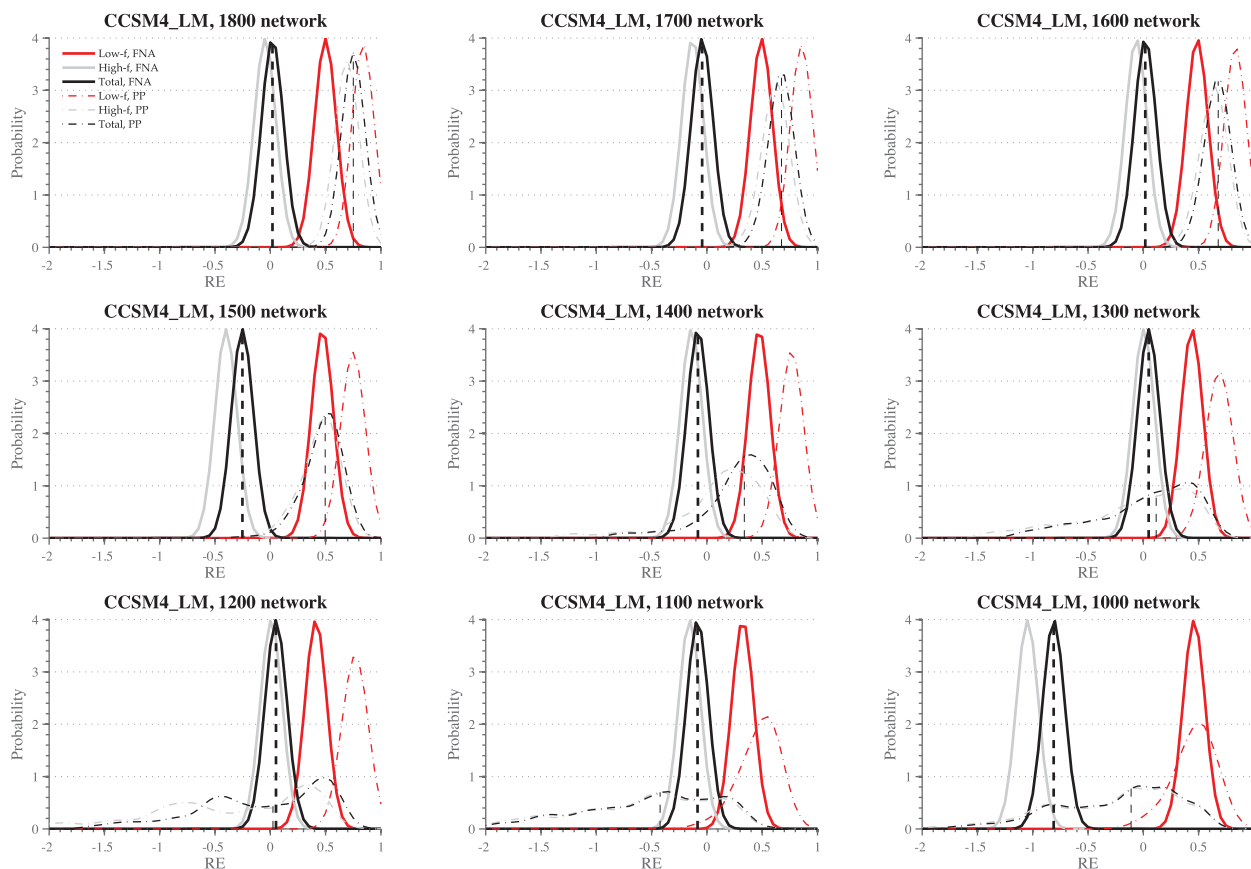


FIG. 13. The FNA Frozen network analysis with LM pseudoproxies in the SNR = 1.0 case. Plots represent distributions obtained by kernel density smoothing of the raw histogram for each century, with a bandwidth of 0.1. Solid lines represent the FNA distributions (short calibration), while dash-dotted lines represent the PP distributions (long calibration). (cf. Fig. 3).

thereby providing an overly pessimistic assessment of the quality of LF reconstructions based on longer training and verification periods. This is cause for concern, because cross-validation exercises of this type are currently the only way to evaluate the reliability of real-world climate reconstructions, and our results suggest that this may be very difficult with inevitably short calibration/verification periods. Furthermore, the frozen network exercise is inherently oblivious to the growth of proxy age uncertainties with time. Because slight offsets between proxies can readily cause destructive interferences between interannual signals, ignoring them will tend to overestimate HF reconstruction skill in preinstrumental times. However, this problem is alleviated at low frequencies given a dating control of 5 yr or better.

We further probed RegEM's performance via a suite of realistic pseudoproxy tests using two different GCM simulations of the past millennium, one driven by best estimates of solar and volcanic forcing (LM) and the other driven by constant forcing (PI). The range of

tropical Pacific variability spanned by these two cases was meant to bracket nature's behavior, and indeed encompasses the behavior of the three instrumental Niño-3.4 time series considered in this study (Fig. 6). Our pseudoproxy design incorporates key aspects of the real network, including the decreasing availability of predictors back in time, and uses pertinent climate fields as input for the pseudoproxies. It does, however, make very stringent assumptions regarding the physical controls of each proxy, sometimes overestimating and sometimes underestimating their relationship to Niño-3.4. The case of most interest (SNR = 1.0) suggests that about 40% of the total Niño-3.4 variance (75% of LF variance) can be reliably reconstructed back in time, consistent with frozen network experiments, although with important variance losses at LF and a warm bias induced by the presence of a strong trend over the calibration interval. The large LF RE scores found in this case very likely constitute an upper bound on the real-world performance of our network and method. Discrepancies between FNA and

PPEs are partially reconciled by considering the inability of FNA to approximate the true PP distributions (section 5c).

Nevertheless, important caveats limit the application of those results to the real world: 1) the spectral characteristics of each GCM's simulated Niño-3.4 (none of which emulate instrumental estimates) and 2) the simplicity of the pseudoproxy design. The first point highlights the necessity of using as realistic a GCM as possible in pseudoproxy tests. Millennial integrations run as part of the PMIP3 project (see <https://pmip3.lsce.ipsl.fr/wiki/doku.php/pmip3:design:lm:final>) are bringing the most comprehensive models to bear on the simulation of the tropical Pacific response to natural forcing over the past millennium, and a similar analysis conducted over a wider ensemble of simulations would put our findings on a stronger footing. The second point highlights the importance of increasing the realism of pseudoproxy tests. To the best of our knowledge, this study is the first attempt to date at incorporating a realistic pattern of temporal availability in the pseudoproxy network; that being said, the univariate, white-noise model used here offers only a simplistic representation of the manifold processes whereby climate signals are incorporated into proxy archives. Should the spectral signature of the proxy genesis process assume a more complex form than our simple model, the reconstruction skill would be correspondingly affected between frequency bands, perhaps nonlinearly so. In addition, this pseudoproxy framework (as in all studies to date) assumes perfect age control, which is obviously an idealization. As in FNA, we expect dating uncertainties to negligibly impact LF skill, but deteriorate HF skill. A more rigorous framework for testing reconstruction methods using pseudoproxies consists of using forward, process-based

proxy models, incorporating physical, chemical, and biological mechanisms of proxy genesis to simulate synthetic proxy records with realistic properties (Anchukaitis et al. 2006; Evans 2007; Cobb et al. 2008; Thompson et al. 2011; Mann et al. 2012). This field is only in its infancy, and will await further development before being operationally used for the validation of climate reconstruction methods.

To what extent do these results carry over to real-world reconstructions? Discrepancies between FNA and PPEs illustrate the difficulty of quantitatively assessing LF reconstruction skill: on the one hand, the pessimistic assessments of FNA are most likely excessive; on the other hand, LM may be too favorable a test case because of the relatively high-amplitude LF signals, which exceed that found in all observational datasets considered here. Overall the results suggest that RegEM TTLS can faithfully capture the phase of LF fluctuations, albeit with an amplitude that will inevitably be damped as proxy series drop out of the network—so our estimates are likely to be on the low end of possible centennial variability. It is noteworthy that verification skill depends critically on the instrumental target, which drives significant differences in reconstructed LF behavior. In Part II, we apply the methodologies described and evaluated herein to obtain a range of plausible histories of Niño-3.4 variability over the past 1000 years, before discussing some dynamical implications of the results.

Acknowledgments. We acknowledge funding from NOAA Grant NA06OAR4310116. The authors thank Scott Rutherford, Emanuele Di Lorenzo, Judith Curry, Kettyah Chaak, Tapio Schneider, Julie Richey, David Black, Kevin Anchukaitis, Paige Green, Glenn Alexander, Athanassios Protopapas, and Toby Ault for providing data, technical help, and support.

APPENDIX

Proxy Dataset

Table A1 contains the characteristics of the proxy database that was obtained from NCDC.

TABLE A1. Tropical proxy database; ρ is the correlation of each series with Niño-3.4 (HadSST2i in this case), and γ is the lag-1 autocorrelation obtained by the method of Allen and Smith (1996). Significant correlations are shown in boldface.

No.	Site	Class	Type	Lat	Lon	Range	ρ	γ	Reference
1	Abraham Reef	Coral	$\delta^{18}\text{O}$	22°S	153°E	1638–1983	+0.03	+0.49	Druffel and Griffin (1999)
2	Abrolhos	Coral	$\delta^{18}\text{O}$	28°S	114°E	1795–1995	+0.04	+0.16	Kuhnert et al. (1999)
3	Amedee Light	Coral	$\delta^{18}\text{O}\delta^{18}\text{O}$	22°S	166°E	1661–1994	+0.03	+0.63	Quinn et al. (1998)
4	Aqaba 18	Coral	$\delta^{18}\text{O}$	30°N	35°E	1788–1992	−0.19	+0.20	Heiss (1994)
5	Bermuda	Coral	Sr/Ca	32°N	64°W	1781–1998	−0.07	+0.70	Goodkin et al. (2008)
6	Espiritu Santo	Coral	$\delta^{18}\text{O}$	15°S	167°E	1806–1979	+0.26	+0.61	Quinn et al. (1996)
7	Fiji 1F	Coral	$\delta^{18}\text{O}$	17°S	179°E	1781–1996	+0.28	+0.70	Linsley et al. (2008)

TABLE A1. (Continued)

No.	Site	Class	Type	Lat	Lon	Range	ρ	γ	Reference
8	Fiji AB	Coral	$\delta^{18}\text{O}$	17°S	179°E	1617–2001	+ 0.31	+0.51	Linsley et al. (2008)
9	Guam	Coral	$\delta^{18}\text{O}$	13°N	145°E	1791–2001	+0.06	+0.73	Asami et al. (2005)
10	Ifaty	Coral	$\delta^{18}\text{O}$	23°S	44°E	1660–1996	+ 0.25	+0.24	Zinke et al. (2004)
11	Ifaty	Coral	Sr/Ca	23°S	44°E	1689–1996	+0.08	+0.03	Zinke et al. (2004)
12	Kavieng	Coral	Sr/Ca	3°S	151°E	1824–1998	+0.10	+0.31	Alibert and Kinsley (2008)
13	La Reunion	Coral	$\delta^{18}\text{O}$	21°S	55°E	1833–1996	−0.09	+0.61	Pfeiffer et al. (2004)
14	Lombok Strait	Coral	$\delta^{18}\text{O}$	8°S	116°E	1783–1991	+ 0.20	+0.31	Charles et al. (2003)
15	Mafia	Coral	$\delta^{18}\text{O}$	8°S	40°E	1623–1999	− 0.42	+0.35	Damassa et al. (2006)
16	Mahe	Coral	$\delta^{18}\text{O}$	5°S	55°E	1847–1996	− 0.42	+0.41	Charles et al. (1997)
17	Maiana Atoll	Coral	$\delta^{18}\text{O}$	1°N	173°E	1841–1995	− 0.66	+0.50	Urban et al. (2000)
18	Malindi	Coral	$\delta^{18}\text{O}$	3°S	40°E	1801–1994	− 0.30	+0.71	Cole et al. (2000)
19	Palmyra	Coral	$\delta^{18}\text{O}$	6°N	162°W	1146–1998	− 0.63	+0.52	Cobb et al. (2003)
20	Rarotonga (2R)	Coral	$\delta^{18}\text{O}$	21°S	160°E	1727–1997	+ 0.26	+0.41	Linsley et al. (2008)
21	Rarotonga (2R)	Coral	Sr/Ca	21°S	160°E	1766–1997	+ 0.34	+0.32	Linsley et al. (2006)
22	Ras Umm Sidd	Coral	$\delta^{18}\text{O}$	28°N	34°E	1752–1996	− 0.15	+0.28	Felis et al. (2000)
23	Savusavu	Coral	$\delta^{18}\text{O}$	17°S	179°E	1776–2001	+ 0.37	+0.65	Bagnato et al. (2005)
24	Secas	Coral	$\delta^{18}\text{O}$	8°N	442°W	1708–1985	+0.00	+0.30	Linsley et al. (1994)
25	Tonga TH1	Coral	$\delta^{18}\text{O}$	20°S	535°W	1794–2004	+ 0.22	+0.64	Linsley et al. (2008)
26	Turumote	Coral	Sr/Ca	18°N	67°W	1751–2004	−0.08	+0.37	Kilbourne et al. (2008)
27	Urvina Bay	Coral	$\delta^{18}\text{O}$	0°	451°W	1607–1982	−0.06	+0.37	Dunbar et al. (1994)
28	Salalah	Speleothem	$\delta^{18}\text{O}$	18°N	55°E	1218–1998	+0.07	+0.70	Burns et al. (2002)
29	Puruogangri, Tibet	Ice core	$\delta^{18}\text{O}$	34°N	89°E	1603–1998	+0.02	+0.99	Thompson et al. (2006b)
30	Huascarán, Peru	Ice core	$\delta^{18}\text{O}$	14°S	71°W	1598–1993	+0.04	+0.97	Thompson et al. (1995)
31	Sajama, Bolivia	Ice core	$\delta^{18}\text{O}$	18°S	69°W	1603–1998	−0.05	+0.96	Thompson et al. (1998)
32	Quelccaya, Peru	Ice core	$\delta^{18}\text{O}$	14°S	71°W	1540–2002	+ 0.37	+0.54	Thompson et al. (2006a)
33	Lake Challa	Sediment	Log(varve thickness)	3°S	38°E	−1050–2005	− 0.54	−0.07	Wolff et al. (2011)
34	Cariaco basin	Sediment	Mg/Ca SST	11°N	65°W	1221–1990	+ 0.29	+0.62	Black et al. (2007)
35	NADA PDSI	Tree	PC1	28°N	105°W	0–2006	− 0.56	+0.29	Cook (2008)
36	Saradan	Tree	TRW	8°S	111°E	1690–2000	+0.02	+0.52	Stahle et al. (1998)
37	Zimbabwe	Tree	Log (summer Pr)	19°S	27°E	1796–1996	− 0.15	+0.00	Therrell et al. (2006)
38	Chilean Cordillera	Tree	Log (summer Pr)	33°S	72°W	1000–2000	+ 0.28	−0.12	Le Quesne et al. (2006)
39	Barranca	Tree	PDSI	19°N	99°W	771–2008	−0.04	+0.44	Stahle et al. (2011)
40	Madera Canyon, Mexico (MX)	Tree	TRW	29°N	103°W	1675–1998	+ 0.38	+0.38	ITRDB (2009)
41	Nevado de Colima, MX	Tree	TRW	20°N	104°W	1553–1998	+0.14	+0.58	ITRDB (2009)
42	Cerro Baraja, MX	Tree	TRW	26°N	106°W	1376–1998	+ 0.46	−0.08	ITRDB (2009)
43	Cerro Baraja, MX	Tree	TRW	26°N	106°W	1376–1998	+ 0.47	−0.05	ITRDB (2009)
44	Creel International Airport, MX	Tree	TRW	28°N	108°W	1644–1998	+ 0.42	−0.06	ITRDB (2009)
45	El Salto, MX	Tree	TRW	24°N	106°W	1481–1998	+ 0.34	+0.36	ITRDB (2009)
46	El Tabacote/Tomochic, MX	Tree	TRW	28°N	108°W	1583–1998	+ 0.41	+0.02	ITRDB (2009)
47	El Vado, MX	Tree	TRW	17°N	97°W	1750–1998	− 0.17	+0.44	ITRDB (2009)
48	Rio Verde, MX	Tree	TRW	22°N	100°W	1574–1998	+ 0.30	+0.38	ITRDB (2009)
49	Villareal, MX	Tree	TRW	20°N	98°W	1710–1998	−0.08	+0.22	ITRDB (2009)
50	Capote Knob, Texas (TX)	Tree	TRW	29°N	98°W	1712–1998	+ 0.18	+0.24	ITRDB (2009)
51	Lavaca River, TX	Tree	TRW	29°N	97°W	1668–1998	+ 0.26	+0.01	ITRDB (2009)
52	Coletto Creek, TX	Tree	TRW	29°N	97°W	1682–1998	+ 0.23	+0.08	ITRDB (2009)
53	Ecleto Creek, TX	Tree	TRW	29°N	98°W	1695–1998	+0.11	+0.39	ITRDB (2009)
54	Camp Springs, TX	Tree	TRW	29°N	103°W	1473–1998	+ 0.40	+0.09	ITRDB (2009)
55	Camp Springs, TX	Tree	TRW	29°N	103°W	1473–1998	+ 0.37	+0.08	ITRDB (2009)
56	Camp Springs, TX	Tree	TRW	29°N	103°W	1473–1998	+ 0.27	−0.01	ITRDB (2009)
57	Die Bos	Tree	TRW	32°S	19°E	1564–1998	−0.08	+0.55	ITRDB (2009)

REFERENCES

- Adams, J., M. Mann, and C. Ammann, 2003: Proxy evidence for an El Niño-like response to volcanic forcing. *Nature*, **426**, 274–278, doi:10.1038/nature02101.
- Alibert, C., and L. Kinsley, 2008: A 170-year Sr/Ca and Ba/Ca coral record from the western Pacific warm pool: 2. A window into variability of the New Ireland Coastal Undercurrent. *J. Geophys. Res.*, **113**, C06006, doi:10.1029/2007JC004263.
- Allen, M. R., and L. A. Smith, 1996: Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise. *J. Climate*, **9**, 3373–3404.
- Ammann, C., and E. Wahl, 2007: The importance of the geophysical context in statistical evaluations of climate reconstruction procedures. *Climatic Change*, **85**, 71–88, doi:10.1007/s10584-007-9276-x.
- Anchukaitis, K. J., M. N. Evans, A. Kaplan, E. A. Vaganov, M. K. Hughes, H. D. Grissino-Mayer, and M. A. Cane, 2006: Forward modeling of regional scale tree-ring patterns in the southeastern United States and the recent influence of summer drought. *Geophys. Res. Lett.*, **33**, L04705, doi:10.1029/2005GL025050.
- Asami, R., T. Yamada, Y. Iryu, T. M. Quinn, C. P. Meyer, and G. Paulay, 2005: Interannual and decadal variability of the western Pacific sea surface condition for the years 1787–2000: Reconstruction based on stable isotope record from a Guam coral. *J. Geophys. Res.*, **110**, C05018, doi:10.1029/2004JC002555.
- Ault, T. R., J. E. Cole, M. N. Evans, H. Barnett, N. J. Abram, A. W. Tudhope, and B. K. Linsley, 2009: Intensified decadal variability in tropical climate during the late 19th century. *Geophys. Res. Lett.*, **36**, L08602, doi:10.1029/2008GL036924.
- Bagnato, S., B. K. Linsley, S. S. Howe, G. M. Wellington, and J. Salinger, 2005: Coral oxygen isotope records of interdecadal climate variations in the South Pacific Convergence Zone region. *Geochem. Geophys. Geosyst.*, **6**, Q06001, doi:10.1029/2004GC000879.
- Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172.
- Black, D. E., M. A. Abahazi, R. C. Thunell, A. Kaplan, E. J. Tappa, and L. C. Peterson, 2007: An 8-century tropical Atlantic SST record from the Cariaco Basin: Baseline variability, twentieth-century warming, and Atlantic hurricane frequency. *Paleoceanography*, **22**, PA4204, doi:10.1029/2007PA001427.
- Box, G., and D. Cox, 1964: An analysis of transformations. *J. Roy. Stat. Soc.*, **26**, 211–252.
- Bradley, R. S., and P. D. Jones, 1993: “Little Ice Age” summer temperature variations: Their nature and relevance to recent global warming trends. *Holocene*, **3**, 367–376, doi:10.1177/095968369300300409.
- Braganza, K., J. L. Gergis, S. B. Power, J. S. Risbey, and A. M. Fowler, 2009: A multiproxy index of the El Niño–Southern Oscillation, A.D. 1525–1982. *J. Geophys. Res.*, **114**, D05106, doi:10.1029/2008JD010896.
- Briffa, K. R., P. D. Jones, J. R. Pilcher, and M. K. Hughes, 1988: Reconstructing summer temperatures in northern Fennoscandia back to A.D. 1700 using tree ring data from Scots pine. *Arct. Alp. Res.*, **20**, 385–394.
- Bürger, G., 2007: On the verification of climate reconstructions. *Climate Past*, **3**, 397–409.
- Burns, S. J., D. Fleitmann, M. Mudelsee, U. Neff, A. Matter, and A. Mangini, 2002: A 780-year annually resolved record of Indian Ocean monsoon precipitation from a speleothem from south Oman. *J. Geophys. Res.*, **107**, 4434, doi:10.1029/2001JD001281.
- Cane, M. A., 2005: The evolution of El Niño, past and future. *Earth Planet. Sci. Lett.*, **230**, 227–240, doi:10.1016/j.epsl.2004.12.003.
- Charles, C. D., D. E. Hunter, and R. G. Fairbanks, 1997: Interaction between the ENSO and the Asian monsoon in a coral record of tropical climate. *Science*, **277**, 925–928, doi:10.1126/science.277.5328.925.
- , K. Cobb, M. D. Moore, and R. G. Fairbanks, 2003: Monsoon–tropical ocean interaction in a network of coral records spanning the 20th century. *Mar. Geol.*, **201**, 207–222.
- Christiansen, B., T. Schmidth, and P. Thejll, 2009: A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness. *J. Climate*, **22**, 951–976.
- Cobb, K. M., C. D. Charles, H. Cheng, and R. L. Edwards, 2003: El Niño/Southern Oscillation and tropical Pacific climate during the last millennium. *Nature*, **424**, 271–276.
- , T. Kiefer, J. M. Lough, J. T. Overpeck, and A. W. Tudhope, 2008: Final report. Representing and reducing uncertainties in high-resolution climate proxy data. Tech Rep., 20 pp. [Available online at <http://www.ncdc.noaa.gov/paleo/reports/trieste2008/trieste2008final.pdf>.]
- Cole, J. E., R. B. Dunbar, T. R. McClanahan, and N. A. Muthiga, 2000: Tropical Pacific forcing of decadal SST variability in the western Indian Ocean over the past two centuries. *Science*, **287**, 617–620, doi:10.1126/science.287.5453.617.
- Collins, M., and Coauthors, 2010: The impact of global warming on the tropical Pacific Ocean and El Niño. *Nat. Geosci.*, **3**, 391–397, doi:10.1038/ngeo868.
- Conroy, J. L., A. Restrepo, J. T. Overpeck, M. Steinitz-Kannan, J. E. Cole, M. B. Bush, and P. A. Colinvaux, 2009: Unprecedented recent warming of surface temperatures in the eastern tropical Pacific Ocean. *Nat. Geosci.*, **2**, 46–50, doi:10.1038/ngeo390.
- Cook, E. R., cited 2008: North American summer PDSI reconstructions, version 2a. NOAA/NGDC Paleoclimatology Program. [Available online at http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD&KeywordPath=Parameters|CLIMATE+INDICATORS|DROUGHT%2FPRECIPIATION+INDICES|PALMER+DROUGHT+SEVERITY+INDEX&EntryId=NOAA_NCDC_PALEO_2004-045&MetadataView=Full&MetadataType=0&Ibnode=mdlb3.]
- , and K. Peters, 1981: The smoothing spline: A new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. *Tree-Ring Bull.*, **41**, 45–53.
- , K. R. Briffa, and P. D. Jones, 1994: Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *Int. J. Climatol.*, **14**, 379–402, doi:10.1002/joc.3370140404.
- Damassa, T. D., J. E. Cole, H. R. Barnett, T. R. Ault, and T. R. McClanahan, 2006: Enhanced multidecadal climate variability in the seventeenth century from coral isotope records in the western Indian Ocean. *Paleoceanography*, **21**, PA2016, doi:10.1029/2005PA001217.
- D’Arrigo, R., E. R. Cook, R. J. Wilson, R. Allan, and M. E. Mann, 2005: On the variability of ENSO over the past six centuries. *Geophys. Res. Lett.*, **32**, L03711, doi:10.1029/2004GL022055.
- , R. Wilson, and A. Tudhope, 2009: The impact of volcanic forcing on tropical temperatures during the past four centuries. *Nat. Geosci.*, **2**, 51–56, doi:10.1038/ngeo393.
- Davis, M., 2001: *Late Victorian Holocausts: El Niño Famines and the Making of the Third World*. Verso, 464 pp.
- Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Stat. Soc.*, **39B**, 1–38.

- Deser, C., A. S. Phillips, and J. W. Hurrell, 2004: Pacific interdecadal climate variability: Linkages between the tropics and North Pacific during boreal winter since 1900. *J. Climate*, **17**, 3109–3124.
- , —, and M. A. Alexander, 2010: Twentieth century tropical sea surface temperature trends revisited. *Geophys. Res. Lett.*, **37**, L10701, doi:10.1029/2010GL043321.
- , and Coauthors, 2012: ENSO and Pacific decadal variability in the Community Climate System Model version 4. *J. Climate*, **25**, 2622–2651.
- DeVries, T. J., and Coauthors, 1997: Determining the early history of El Niño. *Science*, **276**, 965–967, doi:10.1126/science.276.5314.965.
- Dijkstra, H. A., and J. D. Neelin, 1995: Ocean–atmosphere interaction and the tropical climatology. Part II: Why the Pacific cold tongue is in the east. *J. Climate*, **8**, 1343–1359.
- DiNezio, P. N., A. C. Clement, G. A. Vecchi, B. J. Soden, B. P. Kirtman, and S.-K. Lee, 2009: Climate response of the equatorial Pacific to global warming. *J. Climate*, **22**, 4873–4892.
- Druffel, E. R. M., and S. Griffin, 1999: Variability of surface ocean radiocarbon and stable isotopes in the southwestern Pacific. *J. Geophys. Res.*, **104**, 23 607–23 614.
- Dunbar, R. B., G. M. Wellington, M. W. Colgan, and P. W. Glynn, 1994: Eastern Pacific sea surface temperature since 1600 A.D.: The $\delta^{18}\text{O}$ record of climate variability in Galápagos corals. *Paleoceanography*, **9**, 291–316, doi:10.1029/93PA03501.
- Durbin, J., and G. S. Watson, 1951: Testing for serial correlation in least squares regression. II. *Biometrika*, **38**, 159–178, doi:10.1093/biomet/38.1-2.159.
- Emile-Geay, J., 2006: ENSO dynamics and the Earth's climate: From decades to Ice Ages. Ph.D. thesis, Columbia University, 180 pp.
- , K. Cobb, M. Mann, and A. T. Wittenberg, 2013: Estimating central equatorial Pacific SST variability over the past millennium. Part II: Reconstructions and implications. *J. Climate*, **26**, 2329–2352.
- Esper, J., E. R. Cook, and F. H. Schweingruber, 2002: Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science*, **295**, 2250–2253.
- Evans, M. N., 2007: Toward forward modeling for paleoclimatic proxy signal calibration: A case study with oxygen isotopic composition of tropical woods. *Geochem. Geophys. Geosyst.*, **8**, Q07008, doi:10.1029/2006GC001406.
- , A. Kaplan, and M. A. Cane, 1998: Optimal sites for coral-based reconstruction of global sea surface temperature. *Paleoceanography*, **13**, 502–516.
- , —, and —, 2000: Intercomparison of coral oxygen isotope data and historical sea surface temperature (SST): Potential for coral-based SST field reconstructions. *Paleoceanography*, **15**, 551–563.
- , —, and —, 2002: Pacific sea surface temperature field reconstruction from coral $\delta^{18}\text{O}$ data using reduced space objective analysis. *Paleoceanography*, **17** (1), doi:10.1029/2000PA000590.
- Felis, T., J. Pätzold, Y. Loya, M. Fine, A. H. Nawar, and G. Wefer, 2000: A coral oxygen isotope record from the northern Red Sea documenting NAO, ENSO, and North Pacific teleconnections on Middle East climate variability since the year 1750. *Paleoceanography*, **15**, 679–694.
- Fierro, R. D., G. H. Golub, P. C. Hansen, and D. P. O'Leary, 1997: Regularization by truncated total least squares. *SIAM J. Sci. Comput.*, **18**, 1223–1241.
- Frost, C., and S. G. Thompson, 2000: Correcting for regression dilution bias: Comparison of methods for a single predictor variable. *J. Roy. Stat. Soc.*, **163A**, 173–189, doi:10.1111/1467-985X.00164.
- Furtado, J., E. D. Lorenzo, K. Cobb, and A. Bracco, 2009: Paleoclimate reconstructions of tropical sea surface temperatures from precipitation proxies: Methods, uncertainties and non-stationarity. *J. Climate*, **22**, 1104–1123.
- Gagan, M. K., L. K. Ayliffe, J. W. Beck, J. E. Cole, E. R. M. Druffel, R. B. Dunbar, and D. P. Schrag, 2000: New views of tropical paleoclimates from corals. *Quat. Sci. Rev.*, **19**, 45–64, doi:10.1016/S0277-3791(99)00054-2.
- Gergis, J., and A. Fowler, 2009: A history of ENSO events since A.D. 1525: Implications for future climate change. *Climatic Change*, **92**, 343–387, doi:10.1007/s10584-008-9476-z.
- , K. Braganza, A. Fowler, S. Mooney, and J. Risbey, 2006: Reconstructing El Niño–Southern Oscillation (ENSO) from high-resolution palaeoarchives. *J. Quat. Sci.*, **21**, 707–722, doi:10.1002/jqs.1070.
- Goodkin, N. F., K. A. Huguen, W. B. Curry, S. C. Doney, and D. R. Ostermann, 2008: Sea surface temperature and salinity variability at Bermuda during the end of the Little Ice Age. *Paleoceanography*, **23**, PA3203, doi:10.1029/2007PA001532.
- Graham, N. E., 1994: Decadal-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: Observations and model results. *Climate Dyn.*, **10**, 135–162, doi:10.1007/s003820050040.
- , and Coauthors, 2007: Tropical Pacific–mid-latitude teleconnections in medieval times. *Climatic Change*, **83**, 241–285, doi:10.1007/s10584-007-9239-2.
- Hastie, T., R. Tibshirani, and J. Friedman, 2008: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2nd ed. Springer, 745 pp.
- Hegerl, G. C., T. Crowley, M. Allen, W. T. Hyde, H. Pollack, J. Smerdon, and E. Zorita, 2007: Detection of human influence on a new, validated 1500-yr climate reconstruction. *J. Climate*, **20**, 650–666.
- Heiss, G., 1994: Coral reefs in the Red Sea: Growth, production and stable isotopes. Tech. Rep., Leibniz Institute of Marine Sciences, 141 pp.
- Horel, J., and J. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**, 813–829.
- Hoskins, B., and D. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196.
- ITRDB, cited 2009: Tree ring database. International Tree-Ring Data Bank. [Available online at <http://www.ncdc.noaa.gov/paleo/treering.html>.]
- Jones, P., and M. Mann, 2004: Climate over past millennia. *Rev. Geophys.*, **42**, RG2002, doi:10.1029/2003RG000143.
- , and Coauthors, 2009: High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *Holocene*, **19**, 3–49, doi:10.1177/0959683608098952.
- Kanzler, L., 1998: DWATSON: MATLAB module to calculate Durbin–Watson statistic and significance. Tech. Rep., Statistical Software Components, Boston College Department of Economics. [Available online at <http://ideas.repec.org/c/boc/bocode/t850802.html>.]
- Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res.*, **102** (C13), 27 835–27 860.

- , M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.*, **103** (C9), 18 567–18 589.
- , —, and —, 2003: Reduced space approach to the optimal analysis interpolation of historical marine observations: Accomplishments, difficulties, and prospects. Advances in the Applications of Marine Climatology, JCOMM Tech. Rep. 13, WMO, 199–216. [Available online at <http://dev.water.columbia.edu/sitefiles/file/pub/White%20Papers/Cane2003ReducedSpaceApproach.pdf>.]
- Karnauskas, K. B., R. Seager, A. Kaplan, Y. Kushnir, and M. A. Cane, 2009: Observed strengthening of the zonal sea surface temperature gradient across the equatorial Pacific Ocean. *J. Climate*, **22**, 4316–4321.
- Kilbourne, K. H., T. M. Quinn, R. Webb, T. Guilderson, J. Nyberg, and A. Winter, 2008: Paleoclimate proxy perspective on Caribbean climate since the year 1751: Evidence of cooler temperatures and multidecadal variability. *Paleoceanography*, **23**, PA3220, doi:10.1029/2008PA001598.
- Kuhnert, H., J. Pätzold, B. Hatcher, K. H. Wyrwoll, A. Eisenhauer, L. B. Collins, Z. R. Zhu, and G. Wefer, 1999: A 200-year coral stable oxygen isotope record from a high-latitude reef off Western Australia. *Coral Reefs*, **18**, 1–12, doi:10.1007/s003380050147.
- Küttel, M., J. Luterbacher, E. Zorita, E. Xoplaki, N. Riedwyl, and H. Wanner, 2007: Testing a European winter surface temperature reconstruction in a surrogate climate. *Geophys. Res. Lett.*, **34**, L07710, doi:10.1029/2006GL027907.
- Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Capotondi, P. J. Lawrence, and H. Teng, 2013: Last millennium climate and its variability in CCSM4. *J. Climate*, **26**, 1085–1111.
- Le Quesne, C., D. W. Stahle, M. K. Cleaveland, M. D. Therrell, J. C. Aravena, and J. Barichivich, 2006: Ancient *Austrocedrus* tree-ring chronologies used to reconstruct central Chile precipitation variability from A.D. 1200 to 2000. *J. Climate*, **19**, 5731–5744.
- Li, J., S.-P. Xie, E. R. Cook, G. Huang, R. D'Arrigo, F. Liu, J. Ma, and X.-T. Zheng, 2011: Interdecadal modulation of El Niño amplitude during the past millennium. *Nat. Climate Change*, **1**, 114–118.
- Linsley, B. K., R. B. Dunbar, G. M. Wellington, and D. A. Mucciarone, 1994: A coral-based reconstruction of inter-tropical convergence zone variability over Central America since 1707. *J. Geophys. Res.*, **99**, 9977–9994.
- , A. Kaplan, Y. Gouriou, J. Salinger, P. B. deMenocal, G. M. Wellington, and S. S. Howe, 2006: Tracking the extent of the South Pacific convergence zone since the early 1600s. *Geochim. Geophys. Res.*, **7**, Q05003, doi:10.1029/2005GC001115.
- , P. Zhang, A. Kaplan, S. S. Howe, and G. M. Wellington, 2008: Interdecadal–decadal climate variability from multicoral oxygen isotope records in the South Pacific convergence zone region since 1650 A.D. *Paleoceanography*, **23**, PA2219, doi:10.1029/2007PA001539.
- Little, R. J. A., and D. B. Rubin, 2002: *Statistical Analysis with Missing Data*. 2nd ed. Wiley, 381 pp.
- Liu, Z., and M. A. Alexander, 2007: Atmospheric bridge, oceanic tunnel, and global climatic teleconnections. *Rev. Geophys.*, **45**, RG2005, doi:10.1029/2005RG000172.
- Mann, M. E., and S. Rutherford, 2002: Climate reconstruction using 'pseudoproxies.' *Geophys. Res. Lett.*, **29**, 2002, doi:10.1029/2001GL014554.
- , E. Gille, R. Bradley, M. Hughes, J. Overpeck, F. Keimig, and W. Gross, 2000: Global temperature patterns in past centuries: An interactive presentation. *Earth Interact.*, **4**. [Available online at <http://EarthInteractions.org>.]
- , S. Rutherford, E. Wahl, and C. Ammann, 2005: Testing the fidelity of methods used in proxy-based reconstructions of past climate. *J. Climate*, **18**, 4097–4107.
- , —, —, and —, 2007a: Reply. *J. Climate*, **20**, 5671–5674.
- , —, —, and —, 2007b: Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.*, **112**, D12109, doi:10.1029/2006JD008272.
- , Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni, 2008: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl. Acad. Sci. USA*, **105**, 13 252–13 257, doi:10.1073/pnas.0805721105.
- , and Coauthors, 2009: Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. *Science*, **326**, 1256–1260, doi:10.1126/science.1177303.
- , J. D. Fuentes, and S. Rutherford, 2012: Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. *Nat. Geosci.*, **5**, 202–205.
- McShane, B. B., and A. J. Wyner, 2011: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Stat.*, **5**, 5–44, doi:10.1214/10-AOAS398.
- Meehl, G. A., W. M. Washington, T. M. L. Wigley, J. M. Arblaster, and A. Dai, 2003: Solar and greenhouse gas forcing and climate response in the twentieth century. *J. Climate*, **16**, 426–444.
- , and Coauthors, 2007: Global climate projections. *Climate Change 2007: The Physical Science Basis*. S. Solomon et al., Eds., Cambridge University Press, 747–845.
- Moy, C., G. Seltzer, D. Rodbell, and D. Anderson, 2002: Variability of El Niño/Southern Oscillation activity at millennial timescales during the Holocene epoch. *Nature*, **420**, 162–165.
- Nash, J., and J. Sutcliffe, 1970: River flow forecasting through conceptual models. Part I—A discussion of principles. *J. Hydrol.*, **10**, 282–290, doi:10.1016/0022-1694(70)90255-6.
- Ortlieb, L., and J. Macharé, 1993: Former El Niño events: Records from western South America. *Global Planet. Change*, **7**, 181–202, doi:10.1016/0921-8181(93)90049-T.
- Pfeiffer, M., O. Timm, W.-C. Dullo, and S. Podlech, 2004: Oceanic forcing of interannual and multidecadal climate variability in the southwestern Indian Ocean: Evidence from a 160 year coral isotopic record (La Réunion, 55°E, 21°S). *Paleoceanography*, **19**, PA4006, doi:10.1029/2003PA000964.
- Quinn, T. M., T. J. Crowley, and F. W. Taylor, 1996: New stable isotope results from a 173-year coral from Espiritu Santo, Vanuatu. *Geophys. Res. Lett.*, **23**, 3413–3416.
- , —, —, C. Henin, P. Joannot, and Y. Join, 1998: A multicentury stable isotope record from a New Caledonia coral: Interannual and decadal sea surface temperature variability in the southwest Pacific since 1657 A.D. *Paleoceanography*, **13**, 412–426, doi:10.1029/98PA00401.
- Quinn, W. H., 1993: Large-scale ENSO events, the El Niño and other important regional features. *Bull. Inst. Fr. Etud. Andines*, **22**, 13–22.
- Rasmussen, E., and T. Carpenter, 1982: Variations in tropical sea surface temperature and surface winds associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.*, **110**, 354–384.
- Rayner, N., P. Brohan, D. Parker, C. Folland, J. Kennedy, M. Vanicek, T. Ansell, and S. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured

- in situ since the mid-nineteenth century: The HadSST2 data set. *J. Climate*, **19**, 446–469.
- Riedwyl, N., M. Küttel, J. Luterbacher, and H. Wanner, 2008: Comparison of climate field reconstruction techniques: Application to Europe. *Climate Dyn.*, **32**, 381–395, doi:10.1007/s00382-008-0395-5.
- Ropelewski, C., and M. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.*, **115**, 1606–1626.
- Rutherford, S., M. E. Mann, T. L. Delworth, and R. J. Stouffer, 2003: Climate field reconstruction under stationary and non-stationary forcing. *J. Climate*, **16**, 462–479.
- , —, T. J. Osborn, R. S. Bradley, K. R. Briffa, M. K. Hughes, and P. D. Jones, 2005: Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to method, predictor network, target season, and target domain. *J. Climate*, **18**, 2308–2329.
- , —, C. Ammann, and E. Wahl, 2010: Comments on “A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness.” *J. Climate*, **23**, 2832–2838.
- , —, E. R. Wahl, and C. M. Ammann, 2013: Comments on “Erroneous model field representations in multiple pseudoproxy studies: Corrections and implications.” *J. Climate*, in press.
- Sachs, J. P., D. Sachse, R. H. Smittenberg, Z. Zhang, D. S. Battisti, and S. Golubic, 2009: Southward movement of the Pacific intertropical convergence zone AD 1400–1850. *Nat. Geosci.*, **2**, 519–525, doi:10.1038/ngeo554.
- Sardeshmukh, P., and B. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.*, **45**, 1228–1251.
- Schmidt, G. A., and Coauthors, 2012: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1). *Geosci. Model Dev.*, **5**, 185–191, doi:10.5194/gmd-5-185-2012.
- Schneider, T., 2001: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate*, **14**, 853–871.
- , and A. Neumaier, 2001: Algorithm 808: ARfit—A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Software*, **27**, 58–65.
- Seager, R., and G. A. Vecchi, 2010: Greenhouse warming and the 21st century hydroclimate of southwestern North America. *Proc. Natl. Acad. Sci.*, **107**, 21 277–21 282, doi:10.1073/pnas.0910856107.
- , N. Harnik, Y. Kushnir, W. Robinson, and J. Miller, 2003: Mechanisms of hemispherically symmetric climate variability. *J. Climate*, **16**, 2960–2978.
- , —, W. A. Robinson, Y. Kushnir, M. Ting, H. Huang, and J. Velez, 2005: Mechanisms of ENSO-forcing of hemispherically symmetric precipitation variability. *Quart. J. Roy. Meteor. Soc.*, **131**, 1501–1527, doi:10.1256/qj.04.96.
- , A. Tzanova, and J. Nakamura, 2009: Drought in the southeastern United States: Causes, variability over the last millennium, and the potential for future hydroclimate change. *J. Climate*, **22**, 5021–5045.
- Smerdon, J. E., 2012: Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. *WIREs Climate Change*, **3**, 63–77, doi:10.1002/wcc.149.
- , and A. Kaplan, 2007: Comments on “Testing the fidelity of methods used in proxy-based reconstructions of past climate”: The role of the standardization interval. *J. Climate*, **20**, 5666–5670.
- , —, and D. Chang, 2008: On the origin of the standardization sensitivity in RegEM climate field reconstructions. *J. Climate*, **21**, 6710–6723.
- , —, and D. E. Amrhein, 2010a: Erroneous model field representations in multiple pseudoproxy studies: Corrections and implications. *J. Climate*, **23**, 5548–5554.
- , —, D. Chang, and M. N. Evans, 2010b: A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. *J. Climate*, **23**, 4856–4880.
- , —, E. Zorita, J. F. González-Rouco, and M. N. Evans, 2011: Spatial performance of four climate field reconstruction methods targeting the Common Era. *Geophys. Res. Lett.*, **38**, L11705, doi:10.1029/2011GL047372.
- Smith, T., R. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA’s historical merged land–ocean surface temperature analysis (1880–2006). *J. Climate*, **21**, 2283–2296.
- Spiegelhalter, D. J., 1983: Diagnostic tests of distributional shape. *Biometrika*, **70**, 401–409, doi:10.1093/biomet/70.2.401.
- Stahle, D. W., and Coauthors, 1998: Experimental dendroclimatic reconstruction of the Southern Oscillation. *Bull. Amer. Meteor. Soc.*, **79**, 2137–2152.
- , and Coauthors, 2011: Major Mesoamerican droughts of the past millennium. *Geophys. Res. Lett.*, **38**, L05703, doi:10.1029/2010GL046472.
- Sterl, A., G. van Oldenborgh, W. Hazeleger, and G. Burgers, 2007: On the robustness of ENSO teleconnections. *Climate Dyn.*, **29**, 469–485, doi:10.1007/s00382-007-0251-z.
- Therrell, M. D., D. W. Stahle, L. P. Ries, and H. H. Shugart, 2006: Tree-ring reconstructed rainfall variability in Zimbabwe. *Climate Dyn.*, **26**, 677–685, doi:10.1007/s00382-005-0108-2.
- Thompson, D. M., T. R. Ault, M. N. Evans, J. E. Cole, and J. Emile-Geay, 2011: Comparison of observed and simulated tropical climate trends using a forward model of coral of $\delta^{18}\text{O}$. *Geophys. Res. Lett.*, **38**, L14706, doi:10.1029/2011GL048224.
- Thompson, L. G., E. Mosley-Thompson, M. E. Davis, P.-N. Lin, K. A. Henderson, J. Cole-Dai, J. F. Bolzan, and K.-B. Liu, 1995: Late glacial stage and Holocene tropical ice core records from Huascaran, Peru. *Science*, **269**, 46–50, doi:10.1126/science.269.5220.46.
- , and Coauthors, 1998: A 25,000-year tropical climate history from Bolivian ice cores. *Science*, **282**, 1858–1864, doi:10.1126/science.282.5395.1858.
- , and Coauthors, 2006a: Abrupt tropical climate change: Past and present. *Proc. Natl. Acad. Sci.*, **103**, 10 536–10 543, doi:10.1073/pnas.0603900103.
- , E. Mosley-Thompson, M. Davis, T. Mashiotto, K. Henderson, P.-N. Lin, and Y. Tandong, 2006b: Ice core evidence for asynchronous glaciation on the Tibetan Plateau. *Quat. Int.*, **154**, 3–10, doi:10.1016/j.quaint.2006.02.001.
- Thomson, D. J., 1982: Spectrum estimation and harmonic analysis. *Proc. IEEE*, **70**, 1055–1096.
- Tierney, J., D. Oppo, Y. Rosenthal, J. Russell, and B. Linsley, 2010: Coordinated hydrological regimes in the Indo-Pacific region during the past two millennia. *Paleoceanography*, **25**, PA1102, doi:10.1029/2009PA001871.
- Tikhonov, A. N., and V. Y. Arsenin, 1977: *Solution of Ill-Posed Problems*. V. H. Winston and Sons, 258 pp.
- Ting, M., and I. M. Held, 1990: The stationary wave response to a tropical SST anomaly in an idealized GCM. *J. Atmos. Sci.*, **47**, 2546–2566.

- Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam, 2012: Piecing together the past: Statistical insights into paleoclimatic reconstructions. *Quat. Sci. Rev.*, **35**, 1–22, doi:10.1016/j.quascirev.2012.01.012.
- Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2777.
- , G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103**, 14 291–14 324.
- Tudhope, A. W., and Coauthors, 2001: Variability in the El Niño–Southern Oscillation through a glacial–interglacial cycle. *Science*, **291**, 1511–1517, doi:10.1126/science.1057969.
- Urban, F. E., J. E. Cole, and J. T. Overpeck, 2000: Influence of mean climate change on climate variability from a 155-year tropical Pacific coral record. *Nature*, **407**, 989–993.
- Vecchi, G. A., and A. T. Wittenberg, 2010: El Niño and our future climate: Where do we stand? *WIREs Climate Change*, **1**, 260–270.
- Vecchi, G., A. Clement, and B. Soden, 2008: Examining the tropical Pacific's response to global warming. *Eos, Trans. Amer. Geophys. Union*, **89**, 81–83.
- von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, F. González-Rouco, and S. F. B. Tett, 2004: Reconstructing past climate from noisy data. *Science*, **306**, 679–682, doi:10.1126/science.1096109.
- Walker, G. T., and E. W. Bliss, 1932: World weather V. *Mem. Roy. Meteor. Soc.*, **4**, 53–84.
- Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812.
- Weinert, H. L., 2009: A fast compact algorithm for cubic spline smoothing. *Comp. Stat. Data Anal.*, **53**, 932–940, doi:10.1016/j.csda.2008.10.036.
- Whetton, P., and I. Rutherford, 1994: Historical ENSO teleconnections in the eastern hemisphere. *Climatic Change*, **28**, 221–253.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.
- Wilson, R., A. Tudhope, P. Brohan, K. Briffa, T. Osborn, and S. Tett, 2006: Two-hundred-fifty years of reconstructed and modeled tropical temperatures. *J. Geophys. Res.*, **111**, C10007, doi:10.1029/2005JC003188.
- , E. Cook, R. D'Arrigo, N. Riedwyl, M. N. Evans, A. Tudhope, and R. Allan, 2010: Reconstructing ENSO: The influence of method, proxy data, climate forcing, and teleconnections. *J. Quat. Sci.*, **25**, 62–78, doi:10.1002/jqs.1297.
- Wolff, C., G. H. Haug, A. Timmermann, J. S. S. Damsté, A. Brauer, D. M. Sigman, M. A. Cane, and D. Verschuren, 2011: Reduced interannual rainfall variability in East Africa during the last ice age. *Science*, **333**, 743–747, doi:10.1126/science.1203724.
- Worley, S. J., S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott, 2005: ICOADS release 2.1 data and products. *Int. J. Climatol.*, **25**, 823–842, doi:10.1002/joc.1166.
- Yasunaka, S., and K. Hanawa, 2010: Intercomparison of historical sea surface temperature datasets. *Int. J. Climatol.*, **31**, 1056–1073, doi:10.1002/joc.2104.
- Zinke, J., W.-C. Dullo, G. A. Heiss, and A. Eisenhauer, 2004: ENSO and Indian Ocean subtropical dipole variability is recorded in a coral record off southwest Madagascar for the period 1659 to 1995. *Earth Planet. Sci. Lett.*, **228**, 177–194, doi:10.1016/j.epsl.2004.09.028.