Assessing the Significance of Changes in ENSO Amplitude Using Variance Metrics

T. RUSSON, A. W. TUDHOPE, G. C. HEGERL, AND A. SCHURER

School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom

M. COLLINS

College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom

ABSTRACT

The variance of time series records relating to ENSO, such as the interannual anomalies or bandpass filtered components of equatorial Pacific SST indices, provides one approach to quantifying changes in ENSO amplitude. Robust assessment of the significance of changes in amplitude defined in this way is, however, hampered by uncertainty regarding the sampling distributions of such variance metrics within an unforced climate system. The present study shows that the empirical constraints on these sampling distributions provided by a range of unforced CGCM simulations are consistent with the expected parametric form, suggesting that standard parametric testing strategies can be robustly applied, even in the case of the nonlinear ENSO system. Under such an approach, the sampling distribution of unforced relative changes in variance may be constrained by a single parameter τ_d: the value of which depends on the choice of method used to extract the ENSO-related component of time series variability. In the case of interannual anomaly records, the value of τ_d is also substantially dependent on the overall spectral properties of the climatic variable under consideration. In contrast, the τ_d value for bandpass filtered records can be conservatively constrained from the lower edge of the filter passband, allowing for the direct but robust assessment of the significance of relative changes in ENSO amplitude, regardless of the climatic variable under consideration. Example applications of this approach confirm marginally significant F-test p values for multidecadal changes in central Pacific instrumental SST variance and highly significant ones for centennial changes in central Pacific coral δ^{18}O variance.

1. Introduction

The El Niño–Southern Oscillation (ENSO) phenomenon arises in the coupled ocean–atmosphere dynamics of the tropical Pacific and constitutes the most significant mode of interannual climate variability in the modern climate system. Instrumental records of tropical Pacific sea surface temperature (SST) show a decline in the average magnitude of ENSO events from the late nineteenth century into the mid-twentieth century, followed by a larger increase into the late twentieth century (Rayner et al. 2003). Proxy reconstructions of ENSO extend the historical record to cover sizeable portions of the last millennium and show multidecadal to centennial changes in variability as large as those seen over the twentieth century (Cobb et al. 2013; McGregor et al. 2010; Li et al. 2011; Fowler et al. 2012). The amplitude of ENSO variability is thought to be sensitive to the mean-state climate of the tropical Pacific (Cane and Zebiak 1985; Guilyardi 2006; Choi et al. 2012; Watanabe and Wittenberg 2012), such that the historical changes may represent detectable responses to forced changes in mean state. However, long unforced climate model simulations are also capable of yielding relatively large multidecadal changes in the amplitude of ENSO variability (Wittenberg 2009), suggesting that detecting such a response against the range of behavior associated with the unforced ENSO system may not be easy.

To quantitatively address whether historical changes in the amplitude of ENSO variability are significant, in the sense of being unlikely to have arisen from random time-domain sampling within an unforced climate system, it is necessary to both select a metric with which to

Corresponding author address: T. Russon, School of GeoSciences, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JW, United Kingdom.
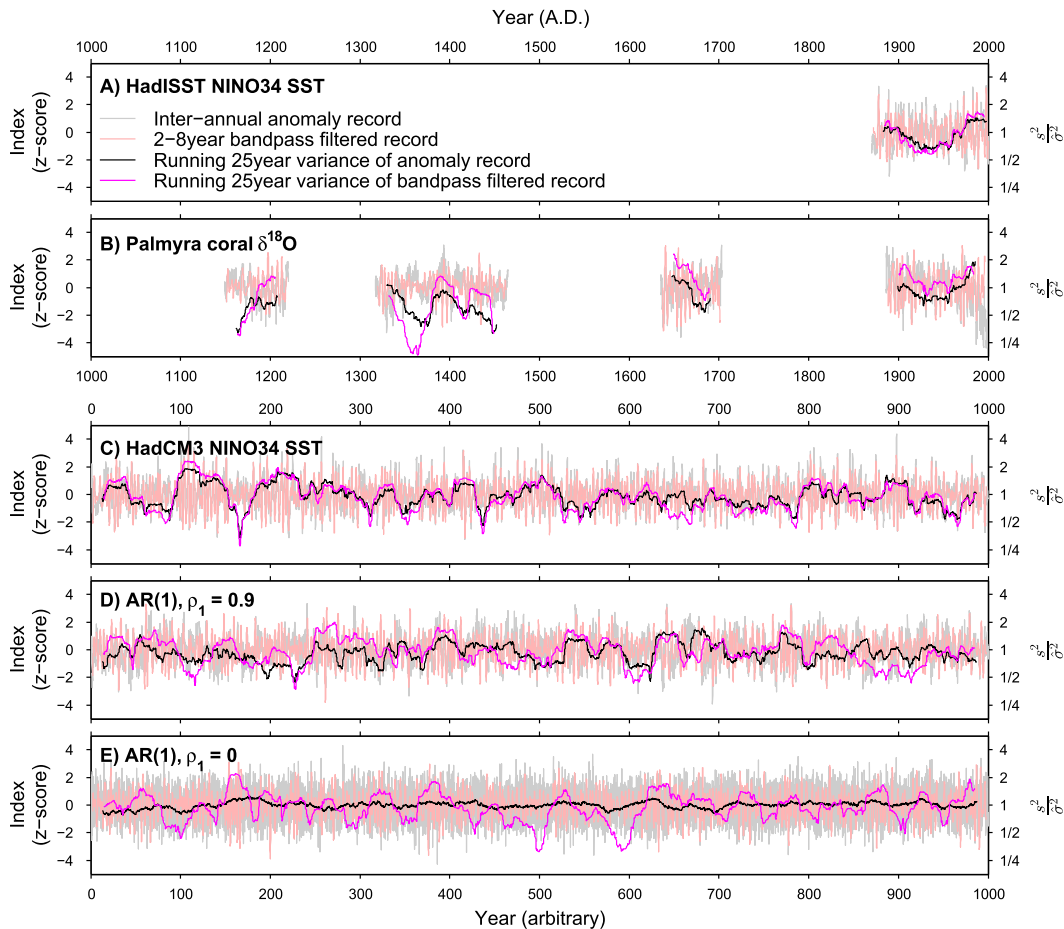E-mail: tom.russon@ed.ac.uk

FIG. 1. Monthly resolution time series and relative changes in multidecadal ENSO-related variance as seen in two realizations of the historical climate system and three realizations of a possible unforced climate system. In each case, interannual anomaly and 2–8-yr Lanczos bandpass filtered records are plotted as z scores (mean subtracted and normalized to unit variance) against the left-hand axes. The bandpass filtered records are slightly shorter than the anomaly ones because of the omission of data equivalent to the edge of the lower passband edge from either end of each record to account for filter edge effects. The 25-yr running variance of the anomaly and bandpass filtered records s^2, normalized in each case to the whole record variance $\hat{\sigma}^2$, are then plotted against the logarithmic right-hand axes. The two historical realizations are (a) Niño-3.4 SST from HadISST (Rayner et al. 2003) and (b) the Palmyra coral $\delta^{18}O$ records (Cobb et al. 2003). In the case of the coral records, the whole record variances used for the relative running variance normalizations are the weighted averages across all the records. The three unforced climate realizations are (c) Niño-3.4 SST from an unforced simulation of the HadCM3 CGCM, (d) an AR(1) process with ρ_1 of 0.9, and (e) white noise [equivalent to AR(1) with $\rho_1 = 0$]. We do not show the ENSO-related components of HadCM3 Niño-3.4 pseudocoral variability as these are extremely similar in this case to those seen for Niño-3.4 SST alone.

quantify ENSO amplitude over any specified interval and to define the unforced sampling distribution of that metric. Such metrics may be based on the properties of individual events (Hereid et al. 2013) or on the variance of time series that are thought to be dominated by ENSO-related variability. The latter approach, which forms the focus of the present study, is widely used in paleoclimatic studies, due principally to the difficulty of objectively identifying individual events in noisy data. To illustrate the variance approach, Fig. 1a shows the 25-yr running

variance of two measures of the ENSO-related variability present within the central equatorial Pacific (5°N–5°S and 170°–120°W) Niño-3.4 SST index, as calculated from the monthly resolution instrumental Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST; Rayner et al. 2003). The two methods used to extract the ENSO-related component of variability are the removal of the averaged annual cycle to leave an interannual anomaly record and the application of a 2–8-yr bandpass filter. As an example of an unforced CGCM simulation containing

qualitatively similar time-domain changes in ENSO amplitude to those seen in the historical records, Fig. 1c shows the running variances of the same two measures of ENSO-related Niño-3.4 SST variability derived from a 1000-yr preindustrial control simulation of the Met Office Hadley Centre Coupled Model, version 3 (HadCM3), coupled general circulation model (CGCM) (Schurer et al. 2014; Gordon et al. 2000; Collins et al. 2001). The Palmyra coral δ^{18}O records presented in Cobb et al. (2003) are also shown as an example of a set of last millennium central Pacific proxy records that are of monthly resolution and capture the first-order multidecadal changes in twentieth-century ENSO amplitude seen within the instrumental Niño-3.4 SST index (Figs. 1a,b).

In spite of the fact that sample variance is a well-defined statistical property for which the expected parametric form of the sampling distribution is well known (von Storch and Zwiers 2003), there are two major difficulties associated with establishing the unforced sampling distributions of ENSO-related variance metrics. First, nonlinear aspects of ENSO dynamics lead to central and eastern Pacific SST indices being positively skewed (An et al. 2005) and the potential for heteroscedastic unforced behavior (Cane et al. 1995; Timmermann 2003; Choi et al. 2012), such that the parametric form may not be appropriate. Second, method-specific constraints will be needed, as the sampling distributions should be expected to differ according to the choice of method used to define the extract the ENSO-related component of variability from the original time series record. Such method dependency is illustrated in the differences between the running variance plots of the interannual anomaly and bandpass filtered records in Figs. 1a–c, which are in turn seen to vary as a function of the climatic variable under consideration. Taken collectively, the two factors described above make the practical assessment of the significance of historical changes in ENSO amplitude using variance metrics challenging.

The motivation for the present study is to investigate possible approaches to assessing the significance of historical changes in ENSO amplitude using variance metrics. In particular, we are interested in whether any simple to use but robust approaches to making such assessments exist. The first section of the study considers whether the expected parametric sampling distribution form is likely to provide an adequate description for unforced changes in the variance of ENSO-related records, based on the behavior seen in CGCM simulations. The outcome of this analysis is used to determine whether standard parametric tests can be employed to ENSO-related variance or whether application-specific nonparametric approaches are likely to be necessary. As it is impractical to evaluate

all possible records of ENSO-related variability, results are presented in the first instance for the example cases of the interannual anomalies and bandpass filtered components of monthly resolution Niño-3.4 SST indices. Sensitivity tests to the use of eastern and western equatorial Pacific indices are also discussed where appropriate. To facilitate comparison with the bivariate δ^{18}O proxy system, which responds to the oxygen isotopic composition of the seawater as well as SST, the analysis was also performed using CGCM-derived pseudocoral indices. As most CGCMs do not contain water isotopic variables, we follow the salinity-based approach of Thompson et al. (2011) to calculate pseudocoral indices. This approach assumes that fluctuations in the isotopic composition of seawater are linearly related to those in salinity, with a slope that can be estimated empirically from instrumental data, such that coral δ^{18}O can also be modeled as a linear combination of SST and salinity. The second section of the study then considers whether absolute or only relative changes in ENSO-related variance can be robustly assessed and whether useful a priori constraints exist on the unforced sampling distributions of variance in interannual anomaly and/or bandpass filtered records. The outcome of this analysis is used to determine whether one or both of these approaches to extracting records of ENSO-related variability has the potential to act as the basis of a simple but robust approach to assessing the significance of historical changes in ENSO amplitude using variance metrics.

2. Can unforced changes in the variance of ENSO-related records be adequately represented by the expected parametric form?

a. Parametric and nonparametric representations of sample variance

The expected distribution of sample variance s^2 for any weakly stationary (homoscedastic) process derived from Gaussian white noise over k data values is given by the χ^2 form [see Eq. (1)] (Knight 2000). The population variance term σ^2 can be thought of here as the long-term average value of unforced variance for the measure of ENSO-related variability under consideration. The other parameter τ_d is a dimensionless factor by which the effective degrees of freedom are reduced relative to the number of original data values. Larger values of τ_d term can be thought of as providing "stricter" constraints on the unforced climate system, as they imply longer decorrelation times and wider χ^2 distributions. Such distributions represent greater unforced modulations in variance, against which more extreme changes would need to be seen in order to be considered significant,
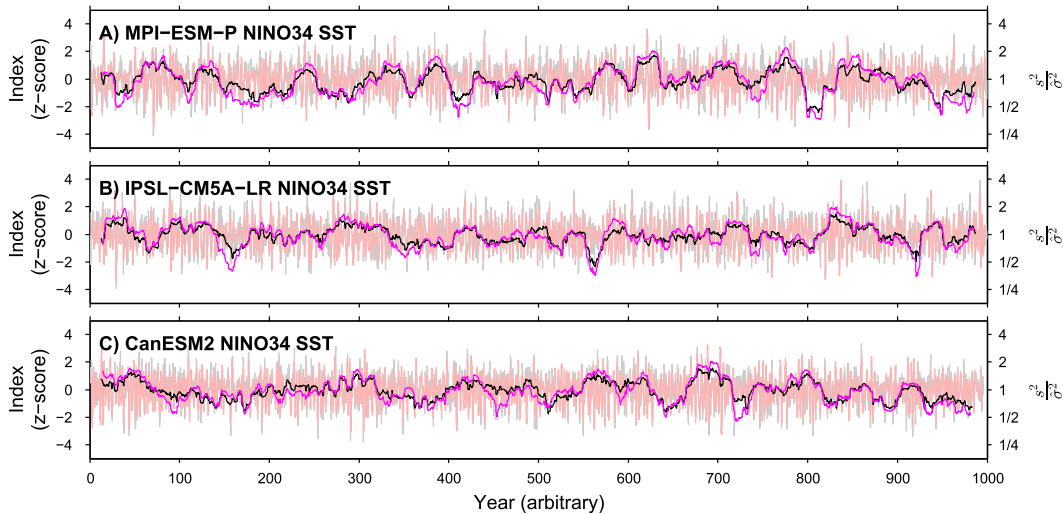
FIG. 2. Monthly resolution time series and relative changes in multidecadal ENSO-related variance as seen in the Niño-3.4 SST indices derived from unforced 1000-yr simulations of the (a) MPI-ESM-P, (b) IPSL-CM5A-LR, and (c) CanESM2 climate models. In each case, interannual anomaly indices and 2–8-yr Lanczos bandpass filtered records are plotted as *z* scores (mean subtracted and normalized to unit variance) against the left-hand axes. The bandpass filtered indices are slightly shorter than the anomaly ones because of the omission of data equivalent to the edge of the lower passband edge from either end of each record to account for filter edge effects. The 25-yr running variance of the anomaly and bandpass filtered records s^2, normalized in each case to the whole record variance $\hat{\sigma}^2$, are then plotted against the logarithmic right-hand axes. The legend for these plots is identical to that in Fig. 1a.

$$s^2 \sim \frac{\sigma^2}{\frac{k}{\tau_d} - 1} \chi^2_{[(k/\tau_d)-1]}. \tag{1}$$

An alternative, nonparametric estimate of the sampling distribution of unforced ENSO-related variance metrics, requiring only the assumption of ergodicity, is given by the empirical cumulative distribution function (ECDF) of all possible sample variances at a given interval length within an unforced time series realization of such a record. One set of nonparametric constraints on the unforced ENSO system that might be expected to be useful are those derived from long, unforced simulations with CGCMs that manifest ENSO-like behavior (Wittenberg 2009; Cobb et al. 2013). In the present study, we analyze interannual anomaly and bandpass filtered records of ENSO-related variability derived from 1000-yr-duration unforced preindustrial simulations taken from different CGCMs. The bandpass filtered records used are derived using a 2–8-yr Lanczos filter design, with twice as many filter weights as data points in the passband (Duchon 1979). The CGCMs were selected based on the criteria of available simulations with that duration of data for both surface ocean temperature and salinity fields and a broad interannual peak in their equatorial Pacific SST power spectrum. The four models are the HadCM3 simulation (as shown in Fig. 1c) and then

three other preindustrial control simulations from the CMIP5 archive: namely, the Max Planck Institute Earth System Model, paleo (MPI-ESM-P); L'Institut Pierre-Simon Laplace Coupled Model, version 5A, low resolution (IPSL-CM5A-LR); and the Canadian Centre for Climate Modelling and Analysis Second Generation Canadian Earth System Model (CanESM2). The interannual anomaly and bandpass filtered records of ENSO-related Niño-3.4 SST variability for the latter three models are shown in Fig. 2. The intercomparison of results from these four different models then provides a semiquantitative way to address the structural uncertainty as to whether any given unforced CGCM simulation useful represents a hypothetical unforced real-world climate. The nonparametric approach does not require either the assumptions of linearity and weak stationarity required by the parametric approach, which may not be justified in the case of the ENSO system (An et al. 2005; Cane et al. 1995; Timmermann 2003; Choi et al. 2012), or the need to estimate the values of the parameters σ^2 and τ_d. However, the finite durations of any practically available realization of unforced ENSO behavior, such as the CGCM control simulations considered here, necessarily means that the associated ECDFs of variance values are incomplete realizations of the true cumulative distribution function (CDF) associated with each models unforced climate system.
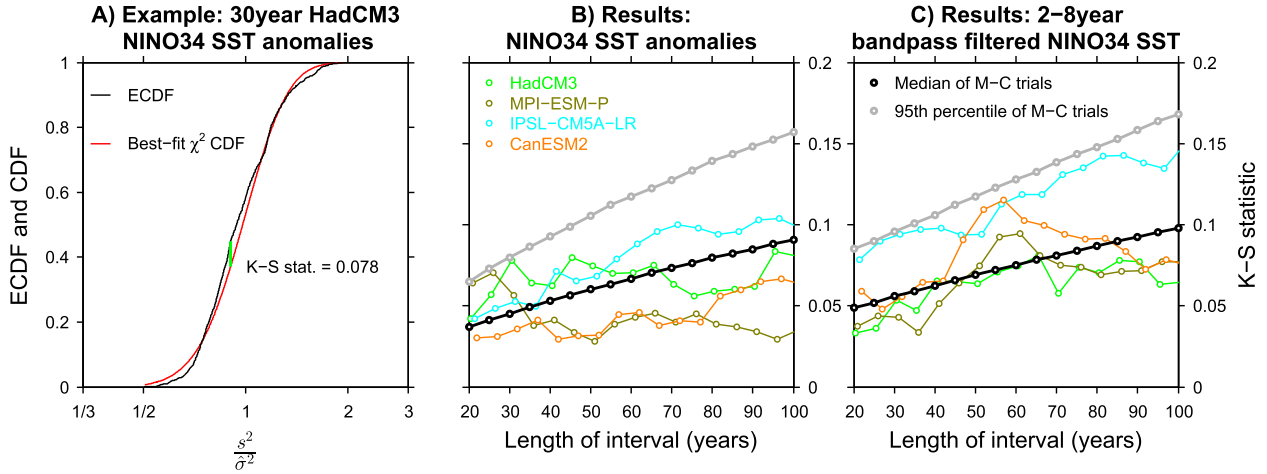
FIG. 3. Illustration of the principle and results of the fit-error method used to evaluate whether the CGCM behavior is consistent with a parametric model for unforced ENSO-related variance. (a) The definition of the K-S statistic as a metric of the fit error between the CGCM-derived ECDF of ENSO-related variance values at a specific interval length and the χ^2 CDF defined by the $\hat{\tau}_d$ and $\hat{\sigma}^2$ estimates obtained from the same data. The x axis is normalized to $\hat{\sigma}^2$ for consistency with the axes in Fig. 1. Also shown are the K-S statistics calculated over interval lengths in the range 20–100 yr, taken at 5-yr increments, for (b) interannual anomaly and (c) 2–8-yr Lanczos bandpass filtered Niño-3.4 SST records derived from the four 1000-yr-duration CGCM simulations. These values are compared to the median and 95th percentile of the same metrics calculated from 5000 MC trials with a suite of equal-duration raw white noise time series.

b. Comparing parametric and nonparametric representations of ENSO-related variance in the CGCM simulations

To investigate whether the variance of ENSO-related records can be adequately represented by the parametric form of Eq. (1), we compare the ECDFs of ENSO-related Niño-3.4 SST variance seen in the four CGCM simulations with estimates of the associated best-fit parametric CDF of such values, as defined by the best-fit estimates of σ^2 and τ_d (denoted by $\hat{\sigma}^2$ and $\hat{\tau}_d$). An estimate of $\hat{\sigma}^2$ is easily estimated for each ENSO-related record from the variance of the entire available time series. Estimating $\hat{\tau}_d$ value is less straightforward and relies upon the relationship between this term and the autocorrelation function of the data ρ [as shown in Eq. (2)] (von Storch and Zwiers 2003; Trenberth 1984). While ρ is not fully known for the finite-duration CGCM-derived records of ENSO-related variability, an empirical estimate is available from the cross-covariance matrix. The estimate of $\hat{\tau}_d$ used here is then obtained from the summation [in the form of Eq. (2)] of the empirical estimate of ρ over the number of lag intervals equivalent to the first two changes in sign,

$$\tau_d = 1 + 2 \sum_{i=1}^{\infty} \rho_i^2. \qquad (2)$$

The statistical framework used to assess the significance of differences between the ECDFs and best-fit parametric CDFs is that of the nonparametric Kolmogorov–Smirnov

(K-S) test, where the test statistic given by the maximum difference between the ECDF and CDF is used to evaluate the null hypothesis of the former having arisen from random sampling within the latter. Figure 3a illustrates the method using the example of 30-yr HadCM3 Niño-3.4 SST interannual anomaly variance values, for which it is seen that the ECDF largely follows the best-fit parametric form but with some differences: the largest of which defines a K-S test statistic value. The significance of such a value cannot, however, be assessed directly from the Kolmogorov distribution itself, as the parameters of the assumed distribution under the null hypothesis have been estimated from the data under consideration. Consequently, a Monte Carlo (MC) approach is employed, based on a suite of 5000 white noise time series of equal length to the CGCM simulations that have been subjected to exactly the same analysis as the CGCM data. White noise provides the lower bound to the expected width of the distribution of K-S statistic values that should be expected for any linear stochastic time series of that duration.

The CGCM-derived K-S statistic values are then compared to those from the MC white noise trials over interval lengths in the range 20–100 yr, evaluated at 5-yr increments. The analysis is performed over this range of interval lengths in order to assess whether the CGCM simulations manifest relatively more important nonstationary behavior in terms of ENSO-related variance at certain time scales. Regardless of the choice of either interannual anomaly (Fig. 3b) or 2–8-yr Lanczos

bandpass filtered (Fig. 3c) records of ENSO-related Niño-3.4 SST variability, the CGCM-derived K-S statistics are seen to lie entirely below the 95% confidence level defined by the MC trials. Sensitivity tests, using either eastern (Niño-3) and western (Niño-4) equatorial Pacific SST indices or the pseudocoral indices from any of these domains, yield the same conclusion. Of the example Niño-3.4 records considered here, the bandpass filtered IPSL-CM5A-LR data are seen to generally yield the highest K-S values, to the extent that these approach the significance threshold for the example bandpass filter design. It is interesting to note that these values always arise from the variance ECDFs being narrower than those of the best-fit CDFs. In practical terms, this means that the IPSL-CM5A-LR simulation exhibits an ENSO phenomenon that is relatively stable, in terms of multidecadal variance, compared to the best-fit stochastic system. Consequently, even if the null hypothesis of that CGCM-derived ECDF having arisen within such a system were to be rejected, the parametric model would still provide a conservative bound for the assessment of the significance of historical changes.

At least for the four CGCM simulations, three spatial domains, and two methods for extracting ENSO-related variability considered here, the null hypothesis of unforced multidecadal changes in ENSO-related variance being drawn from suitably constrained parametric sampling distributions cannot be easily rejected. This suggests that stochastic sampling variability dominates the unforced sampling distributions within these simulations, even though these also resolve at least some key features of the real-world, nonlinear ENSO phenomenon. With the necessary caveat that thee four models may systematically underestimate the nonlinear contribution, this then leads us to propose that the parametric form provides an adequate description of the unforced sampling distributions of variance within ENSO-related records. An alternative expression of this conclusion is that the use of nonparametric statistical tests based on the finite model simulation durations considered here would not yield robustly different outcomes to those derived from parametric ones. Given the much greater ease of generalization associated with the parametric approach, the remainder of the study now focuses on the application of such constraints to assessing past changes in ENSO amplitude.

3. Assessing the significance of changes in ENSO amplitude using variance metrics

a. Whether to assess absolute and/or relative changes in variance

Within the parametric framework outlined in the previous section, the significance of any absolute value of ENSO-related variance can be assessed provided that the two parameters present in Eq. (1) may be constrained to an acceptable level of uncertainty. In an alternative application of the parametric approach, the significance of any relative change in variance between two such records (or independent intervals within a single record) can be assessed by comparing the ratio of two sample variances to the F distribution with degrees of freedom given by $k_1/\tau_d - 1$ and $k_2/\tau_d - 1$, where k_1 and k_2 are the two interval lengths. To evaluate whether the uncertainties in σ^2 and τ_d are likely to allow for the robust assessment of the significance of such absolute and/or relative changes in ENSO amplitude, we now consider the ranges of best-fit values seen for the CGCM records as providing lower bounds to the total uncertainties associated with each term.

The large degree of CGCM dependency in terms of unforced ENSO amplitude (Guilyardi 2006) entails large uncertainties in $\hat{\sigma}^2$. Using Niño-3.4 SST anomalies as an example record of ENSO-related variability, the estimates of $\hat{\sigma}^2$ derived from the four CGCM simulations considered here span a range of almost a factor of 2. Such differences are two orders of magnitude greater than the 5%–95% bootstrapped (10 000 realizations of a moving-block bootstrap with any block length in the range 5–20 yr; Politis and Romano 1994) confidence ranges for each value. Such a large degree of model dependency necessarily introduces unacceptably large uncertainties into the assessment of the significance of absolute values of multidecadal to centennial ENSO-related variance. For example, the level of twentieth-century instrumental Niño-3.4 SST anomaly variance is lower than any such centennial interval within the HadCM3 and CanESM2 simulations, higher than any within the ISPL-CM5 (and specifically the IPSL-CM5A) simulation and roughly central to those seen in the MPI-ESM-P simulation.

The $\hat{\tau}_d$ values derived from the CGCM Niño-3.4 SST anomaly and 2–8-yr Lanczos bandpass filtered records range from 8.3 to 10.3 and from 16.5 to 19.8, respectively. The maximum uncertainty in F-test p value arising from changes in the value of τ_d is not intuitively obvious but can be easily calculated. Relative uncertainties in τ_d of 10%, 35%, and 85% are seen to lead to maximum p value uncertainties of 0.01, 0.05, and 0.1, respectively. The absolute value of τ_d is relatively unimportant for the range of values and interval lengths considered here. Consequently and in contrast to the absolute variance approach, the ranges of CGCM-derived $\hat{\tau}_d$ values are seen to be sufficiently small as to potentially allow for useful assessments of the significance of relative changes in ENSO amplitude.

b. A priori strategies for constraining τ_d

A parametric model of the sampling distribution for unforced relative changes in the variance of ENSO-related records may be constrained by the single parameter τ_d. Unfortunately, τ_d is a nonintuitive term that cannot generally be easily estimated from physical concerns. In particular, while τ_d may be thought of as a form of decorrelation time, it is not equivalent to that which would be used to estimate reductions in degrees of freedom for parametric tests of the mean (von Storch and Zwiers 2003). Furthermore, the value of τ_d is first-order dependent on the method used to extract the ENSO-related component of time series variability, as illustrated by the $\hat{\tau}_d$ values derived from the CGCM bandpass filtered Niño-3.4 records being around 100% greater than those seen for the associated interannual anomaly records. These CGCM-derived $\hat{\tau}_d$ values do themselves provide one set of record-specific constraints on this term but retain the practical limitations of being nontrivial to calculate and model dependent to at least a second-order degree. For these practical reasons, we now consider whether any robust a priori constraints on the τ_d values associated with interannual anomaly and/or bandpass filtered records are available, such as could then form the basis for a simpler approach to assessing the significance of relative changes in ENSO amplitude.

The variance of interannual anomaly records contains contributions from the subannual and decadal or longer components of the original power spectrum, as well as from within the interannual ENSO band. This raises the possibility that such records may be represented, to at least some extent, by simple linear stochastic models for which the autocorrelation structure is fully known. First-order autoregressive processes [AR(1)] provide the simplest category of such models, with a dependence of τ_d value on the lag-1 autocorrelation ρ_1 as shown by

$$\tau_d = \frac{1 + \rho_1^2}{1 - \rho_1^2}. \tag{3}$$

The strategy used here for fitting AR(1) models is based on the assumption that the variability present in the subannual and decadal or longer bands is unforced. The method first requires that the fitted AR(1) process matches the combined variance contained in those parts of the power spectrum, before the best-fit ρ_1 value is determined as that which best captures the ratio of variance present in the subannual and decadal or longer bands. No such AR(1) process can capture the midspectrum peak associated with ENSO, but may instead provide a crude model of the stochastic properties of an "ENSO free" unforced climate system with similar overall spectral properties to the data under consideration (Mann and Lees 1996). The τ_d values estimated from such an approach would, therefore, be expected to represent lower bounds to the $\hat{\tau}_d$ values determined from the unforced CGCM interannual anomaly records, as the latter represents an unforced stochastic system that also contains ENSO variability. For example, the CGCM Niño-3.4 indices, which provide a better match to the overall power spectrum of the instrumental Niño-3.4 SST index than the fitted AR(1) process (Fig. 4a), offer $\hat{\tau}_d$ constraints that are indeed stricter than that from the AR(1) approach (Fig. 5a). Consequently, the use of CGCM-based unforced climate scenarios, rather than the fitting of AR(1) models, has the potential to reduce the inferred significance of historical changes in relative ENSO amplitude seen within instrumental Niño-3.4 interannual anomaly records. However, the changes in significance that arise from such a shift in approach tends to be quite modest, because of the relatively weak dependence of F-test p values to changes in τ_d. For example, the approximate doubling of instrumental Niño-3.4 SST anomaly variance between the two 20-yr intervals of 1920–40 and 1980–2000 yields a one-sided F-test p value of 0.04 against the AR(1)-derived τ_d constraint on an ENSO-free climate. The use of the $\hat{\tau}_d$ constraint from the MPI-ESM-P control simulation would increase the inferred p value to 0.10. That this difference is not larger illustrates that even very simple linear stochastic systems that exclude the dominant ENSO spectral peak exhibit relatively wide sampling distributions for variance, as can be visualized in the running 25-yr variance of an AR(1) process with ρ_1 of 0.9 displaying qualitatively similar changes to those seen in the historical and climate model anomaly records (Fig. 1d).

The extent to which the CGCM-derived $\hat{\tau}_d$ constraints are stricter than those provided by fitted AR(1) models is also dependent on the relative importance of ENSO-related interannual SST variability, relative to the variability present in other parts of the power spectrum. Sensitivity tests with Niño-3 and Niño-4 SST indices show that the relative difference between the CGCM and AR(1) approaches increases as one moves eastward across the equatorial Pacific, consistent with the increasing relative importance of ENSO-related interannual SST variability. As a more extreme example, if the pronounced decadal or longer variability in the Palmyra coral δ^{18}O records were assumed to be unforced in origin, then an AR(1) model fitted to the longest preindustrial proxy record would provide both a better fit to the overall power spectrum than the CGCM-derived Niño-3.4 pseudocorals (Fig. 4b) and a stricter constraint on τ_d than any of the CGCM-derived $\hat{\tau}_d$ values (Fig. 5a). Such a result may well rise because of inadequacies in the CGCM realizations of the equatorial Pacific hydrological cycle
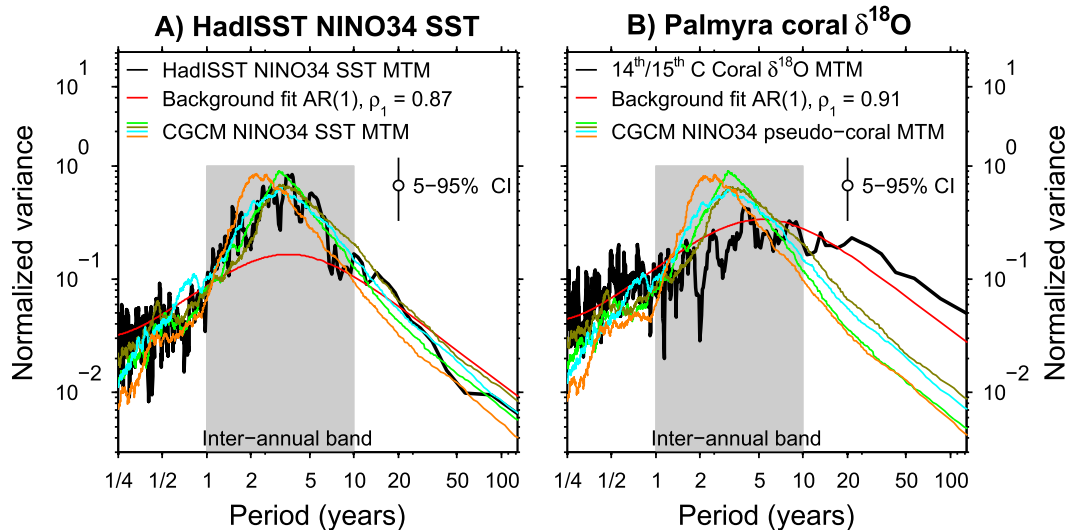
FIG. 4. Multitaper method (MTM) power spectra of (a) the interannual anomalies of the HadISST Niño-3.4 SST data (Rayner et al. 2003) and (b) the fourteenth- and fifteenth-century Palmyra coral δ^{18}O data (Cobb et al. 2003). In both cases, the historical data power spectrum is compared to that of the fitted AR(1) model for that data and also the smoothed MTM spectrum of (a) the Niño-3.4 SST or (b) Niño-3.4 pseudocoral (Thompson et al. 2011) indices taken from the four CGCM simulations. The color coding of the CGCM spectra is the same as that used in Fig. 3. Smoothing of the CGCM spectra is achieved by using a higher time–bandwidth product (64) than that used for the historical data (4). The y axes are given in units of power spectral density multiplied by frequency and divided by total interannual variance, which is equivalent to the normalized fraction of interannual variance present at each frequency. Both the x and y axes are logarithmic. The interannual band (1–10 yr), which is excluded from the fitting of the AR(1) models, is highlighted using background shading. The 5%–95% confidence interval associated with the historical data MTM spectrum is shown as a vertical bar on each plot. The fourteenth- and fifteenth-century coral record is used rather than the twentieth-century one, as the latter yields an even "redder" decadal or longer power spectrum, likely because of the presence of forced hydrological cycle variability. Sensitivity tests show that the ρ_1 values of the fitted AR(1) models are robust to within ± 0.01 for all values of the time–bandwidth product in the range of 2–16.

(Russon et al. 2013; Thompson et al. 2011), leading to the unforced CGCM pseudocorals being too heavily dominated by SST variability, as reflected in the very similar power spectra (Fig. 4) and $\hat{\tau}_d$ values (Fig. 5) seen for the pseudocoral and SST-only Niño-3.4 indices. However, regardless of whether the pronounced decadal or longer variability in the coral records is forced or unforced in origin, the variance of interannual anomaly records taken from such data will contain substantial contamination from variability that is in all likelihood not ENSO related. Applications in which the variance of interannual anomalies are considered a useful method for quantifying ENSO amplitude are therefore typically restricted to climatic variables where the overall power spectrum is strongly dominated by the ENSO peak, such as central and eastern equatorial Pacific SST indices. However, in these situations the AR(1)-derived constraints are shown to be less strict than those derived from unforced CGCM simulations and the extent to which this is the case, along with the absolute values of τ_d, will depend on the particular climatic variable under consideration. Taken collectively, these factors limit the potential for interannual

anomaly records to form the basis of any simple approach to assessing the significance of relative changes in ENSO amplitude using variance metrics.

In contrast to interannual anomalies, the variance of bandpass filtered records is not influenced by subannual and decadal or longer variability, because of the explicit isolation of a defined spectral band. Such an approach reduces the risk of contamination of the variance of such records with non-ENSO-related variability and is hence widely used when considering proxy variables, such as coral δ^{18}O (Tudhope et al. 2001; Cobb et al. 2003, 2013). However, such an approach also necessarily means that the choice of scenario chosen to represent the unforced climate system is of greatly reduced importance, relative to when considering unfiltered records such as interannual anomalies. This reduced dependency arises as the first-order structure of the bandpass filtered power spectrum is determined by the passband itself, and any spectral structure within this as a result of the original overall power spectrum is of secondary importance. To illustrate the extent to which this is the case, the range of τ_d values associated with the entire range of AR(1)
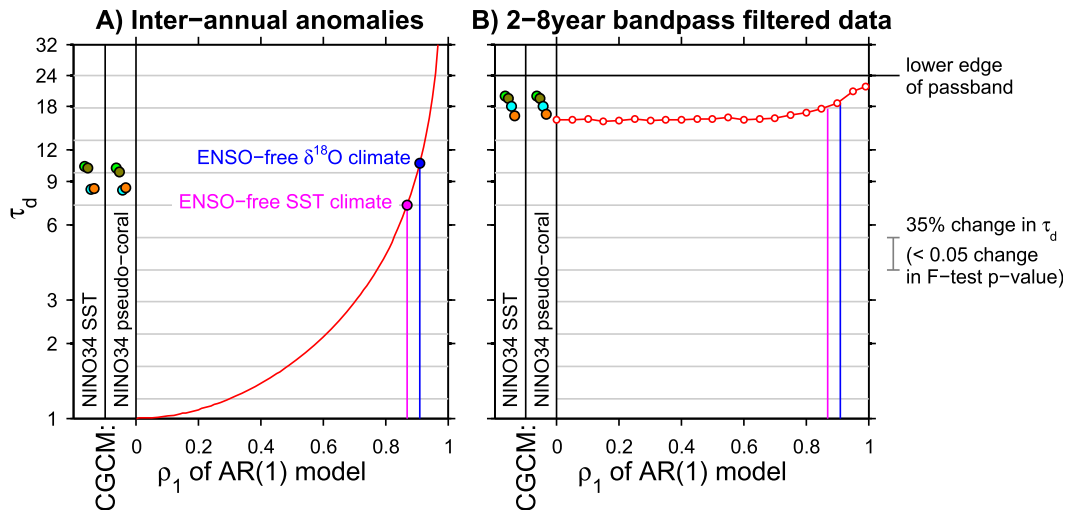
FIG. 5. Comparisons of the constraints available on τ_d from (a) interannual anomaly and (b) 2–8-yr Lanczos bandpass filtered records of ENSO-related variability. In both cases the $\hat{\tau}_d$ values derived from the CGCM Niño-3.4 SST and Niño-3.4 pseudocoral (Thompson et al. 2011) records are shown as the colored circles, with color coding the same as that used in Figs. 3 and 4. Also shown in (a) are the τ_d values associated with unfiltered AR(1) processes, as a function of their ρ_1. For the bandpass filtered records in (b), the $\hat{\tau}_d$ values associated with a representative range of AR(1) processes are determined from 50 000-yr realizations using the method outlined in the text. In both panels, the ρ_1 positions of the AR(1) models fitted to the HadISST Niño-3.4 SST index and fourteenth- and fifteenth-century Palmyra coral δ^{18}O data are highlighted, to allow for comparison of the associated $\hat{\tau}_d$ values with the CGCM-derived ranges. The AR(1)-derived values are labeled as "ENSO free" climate scenarios, as they represent processes with similar overall autocorrelation structures to the historical data but without any midspectral peak in the ENSO band. The τ_d axes are logarithmic, such that a fixed vertical distance represents a fixed relative change in τ_d. The gray reference lines represent a relative change in τ_d of 35%, for which the maximum possible F-test p value uncertainty is 0.05 (provided that the two intervals being compared are of at least 20-yr duration), as shown to the side of (b). Another reference line is also added to (b) to show the proposed bounding τ_d estimate arising from the lower edge of the filter passband.

processes with ρ_1 less than 0.99 is reduced from 1–100 for unfiltered data (Fig. 5a) to 16–24 for the best-fit estimates derived from very long (50 000 yr) 2–8-yr Lanczos bandpass filtered realizations (Fig. 5b). In consequence, the $\hat{\tau}_d$ constraints derived from the bandpass filtered CGCM Niño-3.4 SST records are seen to be of little practical difference, in the context of resultant F-test p values, to those that could have been inferred from the fitted AR(1) models of either the instrumental Niño-3.4 SST or Palmyra coral δ^{18}O data or indeed those available from the bounding case of a white noise ($\rho_1 = 0$) process (Fig. 5b). This result is illustrated in the running relative changes in unfiltered white noise variance being much smaller than those for the other unforced climate scenarios, but this ceases to be the case when the various records are bandpass filtered (Fig. 1e).

The fact that relatively similar bandpass filtered $\hat{\tau}_d$ values can be obtained for a wide range of scenarios for the unforced climate system suggests that a priori constraints based on methodological factors that are applicable across a range of climatic variables may be possible. Empirical models for a lower bound to τ_d, as supplied by

white noise input data, were presented for a range of wavelet filter designs in Torrence and Compo (1998). As it is a conservative upper bound that is sought here, we propose a much cruder rule-of-thumb approach in which a bounding estimate of τ_d is given by the lower edge of the filter passband (i.e., the low-pass filter cutoff), expressed in multiples of the sampling resolution. For example, the bounding τ_d estimate determined in this way for a filter design with the lower edge of the passband being 2 yr and applied to monthly resolution time series data would be 24. That such a bound is conservative to the $\hat{\tau}_d$ values derived from the CGCM Niño-3.4 records is shown for the example 2–8-yr Lanczos filter design in Fig. 4b. If a narrower (wider) passband is used then the value of $\hat{\tau}_d$ increases (decreases), with the lower passband edge remaining the dominant factor, as shown for the examples of the HadCM3 Niño-3.4 SST index and very long realizations of two AR(1) processes in Fig. 6. Across the lower and upper passband edge ranges experimented with here (1.5–2.5 and 7–10 yr, respectively), the lower passband edge is seen to always provide an upper bound to the $\hat{\tau}_d$ values. The choice of filter design does not affect
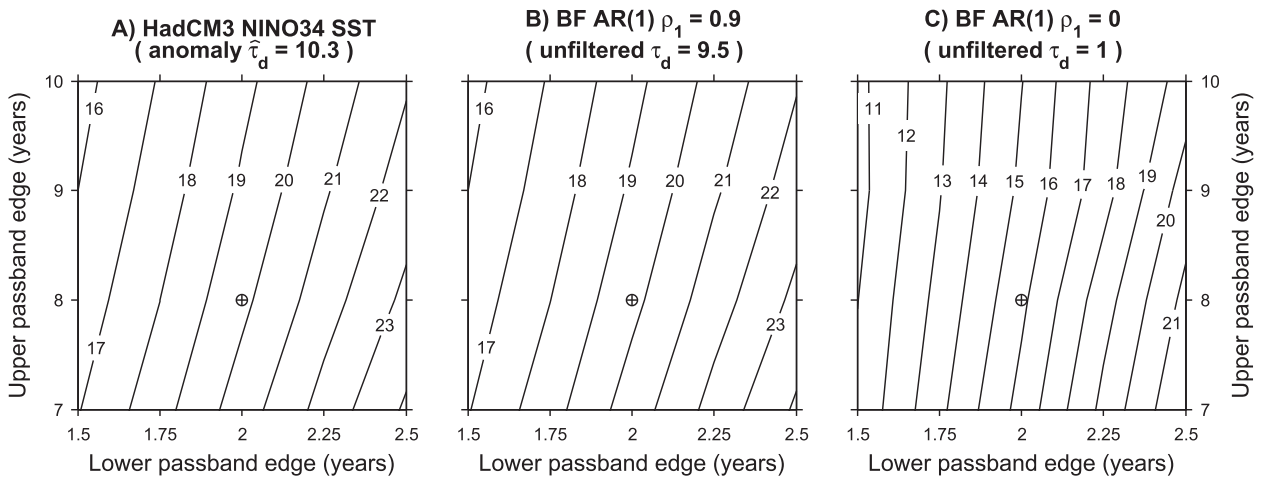
FIG. 6. The sensitivity of the $\hat{\tau}_d$ estimates derived from Lanczos bandpass filtered records to the choice of lower passband edge (low-pass filter cutoff) and upper passband edge (high-pass filter cutoff), as shown for (a) the HadCM3 Niño-3.4 SST index and an AR(1) process with $\rho_1 =$ (b) 0.9 and (c) 0 (white noise). In the latter two cases, the $\hat{\tau}_d$ values are derived using 50 000-yr-long time series. The example passband parameter choice of 2–8 yr, as used throughout the other figures, is highlighted in each panel with crosshairs. The $\hat{\tau}_d$ value associated with the HadCM3 Niño-3.4 SST interannual anomaly record and the analytical ones associated with the unfiltered AR(1) processes are shown underneath the titles, for comparison with the filtered best-fit values. The other three sets of CGCM Niño-3.4 SST data show very similar dependencies to the filter parameters to those seen for HadCM3 (not shown).

the form of the dependency of $\hat{\tau}_d$ on the filter parameters but does exert a second-order effect on the absolute $\hat{\tau}_d$ values. Sensitivity tests show that the $\hat{\tau}_d$ values derived from 2–8-yr bandpass filtered HadCM3 Niño-3.4 SST records using either a fourth-order dual-pass (forward and backward filtering) Butterworth or a fourth-order dual-pass Chebyshev filter design can be up to 15% lower than those for the Lanczos design (not shown). Consequently, the extent to which the lower passband bound is conservative, relative to that of the strictest CGCM-derived constraint, increases with both of the passband parameters and is also affected by the choice of filter design. However, in none of the sensitivity tests undertaken here does this relative difference exceed 52% (37% if only the Lanczos filter design is considered), which would still lead to modest uncertainties in resultant F-test p values. While other filter designs and/or passband choices may lead to situations in which the proposed τ_d bound ceases to be conservative or becomes conservative to an excessive degree, the present sensitivity tests suggests that the proposed bound should be usefully applicable in a wide range of ENSO-related bandpass filtering applications. One necessary limitation to the bandpass filtering approach is that it introduces edge effect uncertainties at least equal to the lower edge of the filter passband at the beginning and end of each discrete record. Accounting for these edge effects slightly reduces the effective duration of the remaining time series and hence the degrees of freedom

present in the data, which may affect the utility of bandpass filtering when considering very short (<10 yr) records.

If the increase in ENSO amplitude seen between 1920–40 and 1980–2000 in the HadISST Niño-3.4 index is now reconsidered using a 2–8-yr Lanczos bandpass filtered record and the bounding τ_d estimate of 24, a similar one-sided F-test p value (0.08) is recovered to that seen for the associated interannual anomaly record against the strictest CGCM-derived constraint. As a second example, which allows for centennial rather than multidecadal record lengths to be compared, we also consider the roughly twofold increase in bandpass filtered variance seen between the preindustrial fourteenth–fifteenth-century and twentieth-century bandpass filtered Palmyra coral $\delta^{18}O$ records. Again using the bounding τ_d estimate, the relative change in sample variance yields a highly significant one-sided F-test p value of less than 0.01. These two examples illustrate the potential of the conservatively bounding rule-of-thumb approach to still yield robustly significant results, in applications to climatic variables with widely differing overall power spectra. It should, however, be noted that in the case of the coral comparison the quoted significance levels apply only to the change in the measured proxy variable ($\delta^{18}O$ at one location) and cannot be translated, without consideration of further uncertainties, to a significance level for any associated change in ENSO-related Niño-3.4 SST variance.

4. Conclusions

This study has investigated the use of variance metrics as measures of ENSO amplitude and how the significance of changes in such metrics, against the unforced variability of the climate system, may be robustly assessed. We propose that the expected parametric forms of the sampling distributions for unforced changes in variance within ENSO-related records are likely to be adequate, based on the behavior seen in a range of unforced CGCM simulations being consistent with such forms. However, inter-CGCM dependency in terms of the long-term unforced amplitude of ENSO means that only relative changes in the variance of ENSO-related records can be robustly assessed. A single parameter is then required to define the sampling distribution associated with such relative changes: the value of which depends to the first order on the method used to extract the ENSO-related component of time series variability. We have considered strategies by which a priori estimates of this term can be made for interannual anomaly and bandpass filtered records. Based on these investigations, we propose a simple and robustly conservative approach to assessing the significance of relative changes in ENSO amplitude based on variance metrics, regardless of the climatic variable under consideration. In this approach, the ratio of sample variance between two independent bandpass filtered intervals is used as the test statistic for the standard parametric F test, with the degrees of freedom constrained by the intervals lengths and the lower edge of the filter passband.

REFERENCES

An, S.-I., Y.-G. Ham, J.-S. Kug, F.-F. Jin, and I.-S. Kang, 2005: El Niño–La Niña asymmetry in the Coupled Model Intercomparison Project simulations. *J. Climate,* **18,** 2617–2627, doi:10.1175/JCLI3433.1.

Cane, M. A., and S. E. Zebiak, 1985: A theory for El Niño and the Southern Oscillation. *Science,* **228,** 1085–1087, doi:10.1126/science.228.4703.1085.

——, ——, and Y. Xue, 1995: Model studies of the long term behaviour of ENSO. *Natural Climate Variability on Decade-to-Century Time Scales,* National Academies Press, 442–456.

Choi, J., S.-I. An, and S.-W. Yeh, 2012: Decadal amplitude modulation of two types of ENSO and its relationship with the mean state. *Climate Dyn.,* **38,** 2631–2644, doi:10.1007/s00382-011-1186-y.

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, doi:10.1038/nature01779.

——, N. Westphal, H. R. Sayani, J. T. Watson, E. D. Lorenzo, H. Cheng, R. L. Edwards, and C. D. Charles, 2013: Highly variable El Niño–Southern Oscillation throughout the Holocene. *Science,* **339,** 67–70, doi:10.1126/science.1228246.

Collins, M., S. F. B. Tett, and C. Cooper, 2001: The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.,* **17,** 61–81, doi:10.1007/s003820000094.

Duchon, C. E., 1979: Lanczos filtering in one and two dimensions. *J. Appl. Meteor.,* **18,** 1016–1022, doi:10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2.

Fowler, A. M., G. Boswijk, A. M. Lorrey, J. Gergis, M. Pirie, S. P. J. McCloskey, J. G. Palmer, and J. Wunder, 2012: Multi-centennial tree-ring record of ENSO-related activity in New Zealand. *Nat. Climate Change,* **2,** 172–176, doi:10.1038/nclimate1374.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.,* **16,** 147–168, doi:10.1007/s003820050010.

Guilyardi, E., 2006: El Niño–mean state–seasonal cycle interactions in a multi-model ensemble. *Climate Dyn.,* **26,** 329–348, doi:10.1007/s00382-005-0084-6.

Hereid, K. A., T. M. Quinn, and Y. M. Okumura, 2013: Assessing spatial variability in El Niño–Southern Oscillation event detection skill using coral geochemistry. *Paleoceanography,* **28,** 14–23, doi:10.1029/2012PA002352.

Knight, K., 2000: *Mathematical Statistics.* Chapman and Hall, 504 pp.

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, doi:10.1038/nclimate1086.

Mann, M. E., and J. M. Lees, 1996: Robust estimation of background noise and signal detection in climatic time series. *Climatic Change,* **33,** 409–445, doi:10.1007/BF00142586.

McGregor, S., A. Timmermann, and O. Timm, 2010: A unified proxy for ENSO and PDO variability since 1650. *Climate Past,* **6,** 1–17, doi:10.5194/cp-6-1-2010.

Politis, D. N., and J. P. Romano, 1994: The stationary bootstrap. *J. Amer. Stat. Assoc.,* **89,** 1303–1313, doi:10.1080/01621459.1994.10476870.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* **108,** 4407, doi:10.1029/2002JD002670.

Russon, T., A. W. Tudhope, G. C. Hegerl, M. Collins, and J. Tindall, 2013: Inter-annual tropical Pacific climate variability in an isotope-enabled CGCM: Implications for interpreting coral stable oxygen isotope records of ENSO. *Climate Past,* **9,** 1543–1557, doi:10.5194/cp-9-1543-2013.

Schurer, A. P., S. F. B. Tett, and G. C. Hegerl, 2014: Small influence of solar variability on climate over the past millennium. *Nat. Geosci.,* **7,** 104–108, doi:10.1038/ngeo2040.

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 δ^{18}O. *Geophys. Res. Lett.,* **38,** L14706, doi:10.1029/2011GL048224.

Timmermann, A., 2003: Decadal ENSO amplitude modulations: A nonlinear paradigm. *Global Planet. Change,* **37** (1–2), 135–156, doi:10.1016/S0921-8181(02)00194-7.

Torrence, C., and G. P. Compo, 1998: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.,* **79,** 61–78, doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.

Trenberth, K. E., 1984: Some effects of sample size and persistence on meteorological statistics. Part I: Autocorrelations. *Mon. Wea. Rev.,* **112,** 2359–2368, doi:10.1175/1520-0493(1984)112<2359:SEOFSS>2.0.CO;2.

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.

von Storch, H., and F. W. Zwiers, 2003: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.

Watanabe, M., and A. T. Wittenberg, 2012: A method for disentangling El Niño–mean state interaction. *Geophys. Res. Lett.,* **39,** L14702, doi:10.1029/2012GL052013.

Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.,* **36,** L12702, doi:10.1029/2009GL038710.