Climate
of the Past

Open Access

Discussions

A probabilistic model of chronological errors in layer-counted climate proxies: applications to annually-banded coral archives

M. Comboul[1], J. Emile-Geay[1], M. N. Evans[2], N. Mirnateghi[1], K. M. Cobb[3], and D. M. Thompson[4]

[1]Department of Earth Sciences, University of Southern California, 3651 Trousdale Parkway, Los Angeles, California, 90089, USA
[2]Department of Geology, University of Maryland, College Park, Maryland, 20742, USA
[3]Department of Earth and Atmospheric Sciences, Georgia Institute of Technology, 311 Ferst Drive Atlanta, Georgia, 30332, USA
[4]Department of Geosciences, University of Arizona, 1040 E. 4th Street Tucson, Arizona, 85721, USA

6077

Abstract

The ability to precisely date climate proxies is central to the reconstruction of past climate variations. To a degree, all climate proxies are affected by age uncertainties, which are seldom quantified. This article proposes a probabilistic age model for proxies based on layer-counted chronologies, and explores its use for annually-banded coral archives. The model considers both missing and doubly-counted growth increments (represented as independent processes), accommodates various assumptions about error rates, and allows to quantify the impact of chronological uncertainties on different diagnostics of variability. In one dimension, we find that time uncertainties primarily affect high-frequency signals but also significantly bias the estimate of decadal signals. We further explore tuning to an independent, tree-ring based chronology as a way to identify an optimal age model. In the multivariate case, a synthetic pseudocoral network is used as testing ground to quantify uncertainties in the estimation of spatiotemporal patterns of variability. Even for small error rates, the amplitude of multidecadal variability is systematically overestimated at the expense of interannual variability (ENSO, in this case), artificially flattening its spectrum at periods longer than 10 yr. An approach to correct chronological errors in coherent multivariate records is presented and validated in idealized cases, though it is found difficult to apply in practice due to the large size of the solution space. We end with a discussion of possible extensions of this model and connections to existing strategies for modeling age uncertainties.

1 Introduction

A peculiar feature of the Earth Sciences is that the independent variable used to peer into Earth's past (the age of the sample) is always uncertain to some degree. In paleoclimatology, time uncertainties depend on the record's structure, which belongs to one of two categories:

1. *Tie-point chronologies*, which include most sedimentary and speleothem archives. Such records are dated via a one-to-one mapping between sample depth and age (that is, an age model). This model typically assumes a piecewise-continuous accumulation rate, constrained by a relatively small number of radiometrically-determined tie points (e.g. [14]C dates for sedimentary archives, or U/Th dates for speleothem archives). Alternatively, when radiometric dating proves impossible, records are often dated by "tuning" to a reference chronology, either derived astronomically (orbital tuning, e.g. Bender, 2002; Hays et al., 1976; Lisiecki and Raymo, 2005; Martinson et al., 1982) or derived from a layer-counted chronology believed to be more accurate (e.g. the GISP2 record Grootes et al., 1993, as done in Wang et al., 2001 and many others).

2. *Layer-counted chronologies*, for instance tree-rings, varved sediments, ice cores and annually-banded corals. For such records, a chronology is obtained by equating the layer count with the number of years or seasons of archive accumulation elapsed since the creation of the topmost (most recent) observed layer. Equivalently, layers may be identified through observation of a seasonally or annually varying quantity such as isotopic composition Fairbanks and Dodge (1979). Layer-counted chronologies are assigned ages relative to a modern tie-point, unless supplemented by independent age controls, such as by radiometric dating (e.g. Cobb et al., 2003), and often assume continuous time resolution unless observations indicate otherwise (e.g. St. George et al., 2013).

Although banded records can often reach subannual precision, they are not time-certain and these uncertainties can have a large influence on variability and coherence of the records. Tree-ring chronologies are generally considered the most accurately dated, but they require adjustments (cross-dating) to optimize coherence across nearby sites (e.g. Douglass, 1941; Fritts, 1976; Wigley et al., 1987; Cook and Kairiukstis, 1990; Yamaguchi, 1991; Stokes and Smiley, 1996). Similarly, although the quoted uncertainty on ice core layer counting is typically on the order of 2 % (Alley et al.,

1997), this value reflects the reproducibility from one chronologist to the next, not the uncertainty about the true date. Seimon (2003) thus showed how two nearby Andean ice cores (Thompson et al., 1985) differed widely when plotted against their published timescales, but that simple adjustments based on volcanic tephra markers could bring the records into excellent agreement. Finally, Emile-Geay et al. (2013) showed how minor age uncertainties in several high-resolution records (in particular coral chronologies) could mute interannual signals in a multiproxy reconstruction of tropical Pacific sea-surface temperature.

Time uncertainties pose a serious challenge to the analysis of paleoclimate signals, as the mathematical machinery devoted to timeseries analysis (e.g. spectral analysis, singular spectrum analysis, wavelet analysis), or multivariate analysis (e.g. principal component analysis, regression analysis, climate field reconstruction) implicitly assumes a perfectly known time axis. In the non-banded case, this problem was recognized early-on and motivated the use of tuning, though this approach is not without dangers (Blaauw, 2012). Recently, several studies have chosen instead to explicitly take time uncertainties into account (Huybers and Wunsch, 2004; Blaauw, 2010; Haam and Huybers, 2010; Blaauw and Christen, 2011; Rhines and Huybers, 2011; Anchukaitis and Tierney, 2012), which has greatly helped quantifying the uncertainties inherent to the interpretation of climate signals in such records.

To our knowledge, banded records have not spurred similar modeling efforts, with the notable exception of Rhines and Huybers (2011), whose model targeted long ice-core chronologies like that of GISP2 (Alley et al., 1997). Their study used a discrete random walk, which assumes that the probability of missing a band, or counting a band twice, is symmetric. However, these probabilities are likely asymmetric for most proxies. Further, the model forced the authors to make some unphysical choices in order to match published estimates of age uncertainties. The model we propose is formulated more intuitively in terms of the probabilities of miscounting a band, and naturally allows for asymmetries in the probability of missing or doubly-counting a band. Our work also

applies the model to multivariate datasets and explores ways to correct for age errors using information shared amongst similar records.

Since dendroclimatology has dealt with layer-counted chronological errors for decades, it would be tempting to draw from their extensive literature for other proxy classes. Unfortunately, extending dendrochronological techniques to other proxy types spanning large spatial scales is problematic, because cross-dating relies on the existence of a strong common signal shared among series. This assumption is well suited to trees within a restricted biome but may no longer be valid for other proxy types with banded age models (ice cores, annually-banded corals, varved sedimentary records), either because too few (if any) replicates are available at a given site, or because one is interested in synchronizing records across a wide geographical area (e.g. a pan-Pacific coral network), where the common signal may be weaker. Nevertheless, this approach was successfully applied to replicated coral records from the reefs offshore of Amédée Island, New Caledonia (DeLong et al., 2007, 2013) and from the Gulf of Mexico (DeLong et al., 2011).

In this article we propose a new probabilistic age model for layer-counted records (Sect. 2.1). We first use this model to quantify the uncertainty associated with spectral estimates from univariate coral series (Sect. 2.2). We then generalize this framework to the multivariate setting (Sect. 3.1), evaluating the impact of chronological uncertainties on principal component analysis derived from a pan-Pacific coral network (Sect. 3.3). Leveraging the mutual information from these signals, we then propose a method to correct for age uncertainties in banded records (Sect. 4), which we validate on synthetic data. Discussion follows in Sect. 5.

2 Banded age model: design and application

2.1 Univariate model formulation

Denote by $X := \{\boldsymbol{x}, \boldsymbol{t}\}$ a banded time series, where $\boldsymbol{x}$ is the sequence of n observations and $\boldsymbol{t}$ its corresponding chronology (i.e. the ith measurement x_i, $i \in \{1, 2, \ldots, n\}$ is assigned a time t_i). The time axis is modeled as follows: starting from the most recent observation (collection date) $t_1 = t_\mathrm{f}$, whose date is assumed to be known exactly, earlier times are deduced by assuming that observations are consecutive, so previous observations $t_i, i \in \{2, 3, \ldots, n\}$ follow the relation:

$$t_i = t_{i-1} - \Delta_i \tag{1}$$

where $\Delta \in \mathbb{N}^{n-1}$ is a vector of time increments. For error-free annually banded proxies, $\Delta_i = \Delta = 1\,\mathrm{yr}$ exactly. Several events, however, may perturb this ideal chronology. For instance, a coral may fail to form a band during a year of particularly poor growing conditions (related to water turbidity, excessive temperature, or other ecological factors) or years may be missing due to sampling artifacts (e.g., core breaks); a tree may fail to form an annual growth increment in an excessively dry or cold year; an ice core may miss a layer due to low-deposition or high wind ablation that year; and a varved sediment core may miss a year because of low productivity in that particular year. Conversely, a single (large) coral band, in a fast growth state, may be associated with two (multiple) consecutive years instead of a single one. Thus, in reality, Δ takes on a structure in which each entry $\Delta_i = 1 + \max(\epsilon_i, -1)$, where $\epsilon_i \in \mathbb{Z}\,\forall i$, although $\epsilon_i = 0$ most of the time. Considering that error rates are typically going to be highly asymmetric with double counting being more common than missing years (e.g. see asymmetric error rates in DeLong et al., 2013), we use a stochastic model (see Appendix A), in which the increment matrix comprises two independent stochastic processes P^{θ_1} and P^{θ_2} with parameters θ_1 and θ_2 respectively such that:

$$\forall i \in [1, 2, \ldots, n-1], \Delta_i = 1 + P_i^{\theta_1} - \min\left(P_i^{\theta_2}, 1\right) \tag{2}$$

and the chronology verifies:

$$t_i = t_\mathrm{f} - i - \sum_{k=1}^{i} P_k^{\theta_1} + \sum_{k=1}^{i} \min\left(P_k^{\theta_2}, 1\right) \tag{3}$$

Age errors are cumulative: for a 100 yr long record, an error rate of 5 % for θ_1 and θ_2 will lead to offsets as large as ±5 yr by the beginning of the sequence, with a probability of ∼ 15 %. Although there is partial compensation between over- and under-counting, leading to a zero *expected* offset in the symmetric case, this expectation is misleading. Indeed, assuming the independence of age errors from one record to another, there is a roughly 2 % chance that the initial date (t_1) of each record will be offset by 10 yr. Though such probability is low, it is non-negligible, and we shall see in Sect. 3.3 that in the case of a multivariate dataset, it may severely bias the estimation of spatiotemporal modes of variability, far beyond the annual scale.

This error model may be used in two ways:

  i. starting from an error-free (synthetic) chronology, one may perturb it by randomly removing or duplicating layers when $P_i^{\theta_1} \geq 1$ or $P_i^{\theta_2} \geq 1$, and linearly reassigning time to the corresponding observations. This allows one to explore the effect of age errors in idealized scenarios.

  ii. Starting from a real chronology, which inevitably contains errors, one may use the model to generate plausible corrections, given some knowledge of the miscounting rate. In this case, extra data entries must be inserted where $P_i^{\theta_1} \geq 1$, while layers must be removed where $P_i^{\theta_2} \geq 1$.

The former case requires interpolation of the data where layers are believed to be missing. While linear interpolation is often used for simplicity, it tends to suppress high frequency variability, hence potentially distorting a record's spectrum. To preserve the intrinsic autocorrelation structure of the timeseries, we use Singular Spectrum Analysis

(SSA, Broomhead and King, 1986; Fraedrich, 1986; Vautard and Ghil, 1989) and its missing value counterpart (SSAM) (Schoellhamer, 2001) to reconstruct the time series including miscounted layers. SSAM was shown to preserve the variance for as many as $\frac{M}{2}$ missing points within a sliding window of size M (Schoellhamer, 2001). In specific applications, other methods would be relevant (e.g. the Lomb-Scargle periodogram Mudelsee et al., 2009, for spectral analysis with missing data) but interpolating with SSAM allows us to apply other analysis methods without further modification, which will prove useful in multivariate applications (Sect. 3).

2.2 Uncertainty quantification in coral records

The above model is now applied to two coral datasets. The first one is the Havannah island fluorescence record (Isdale et al., 1998), chosen for its particularly long, continuous time-span. Because the record already contains dating inaccuracies, we attempt to correct for age errors as explained in (ii). The age uncertainty in coral records is not known, because many of them are not site-replicated. However, DeLong et al. (2013) suggested an error rate of −2.1/+3.7 % in unreplicated samples. We thus assume a conservative error rate of 0.05 for both missing and doubly-counted annual bands (i.e. $\theta_1 = \theta_2 = 0.05$) and generate an ensemble of 1000 plausible age-corrected realizations. The 95 % confidence interval of the realization ensemble is shown along with the original time series in Fig. 1a.

The corresponding spectral densities were estimated in Fig. 1b using the multi-taper method (MTM Thomson, 1982) for each age model realization with a conservative time-bandwidth product of 7/2 (results are insensitive to this choice).

It is generally presumed that dating errors should affect the estimation of interannual variance most strongly, but only negligibly affect decadal signals. Figure 1b shows this expectation to be approximately correct, though the impact on decadal variability is substantial. Indeed, the width of the confidence bands roughly triples from periods of 50 to 5 yr.

It is however noteworthy that the spectral uncertainties due to the age perturbations (shaded area) are markedly narrower than the uncertainties in the spectral estimates themselves (dashed lines, obtained via standard χ^2 intervals), even though a widening of the ensemble spread can be observed as the frequency increases. This result persists with the robust method of Chave et al. (1987) and its associated jackknife intervals.

To check that those results are not an artifact of the choice of coral timeseries, we repeat this analysis on a δ^{18}O record from Palmyra Island (Cobb et al., 2003). Figure 1c shows the cold-season average (December-January-February, DJF) of a continuous splice (1636–1703) of this record, together with the width of the age-perturbed ensemble using $\theta_1 = \theta_2 = 0.05$. Figure 1d shows the impact of age uncertainties on the estimated spectrum, and confirms the previous results: spectral energy of age-corrupt datasets leaks from the interannual band to lower frequencies, but chronological uncertainties generally produce narrower intervals than those due to the spectral estimation process.

2.3 Age model identification

We have shown how the model described in Sect. 2 can be used to generate ensembles of plausible chronologies for any banded record given some a priori estimate of chronological errors. Learning from other dataseries with more replicated and accurate chronologies can then be used to choose an optimal chronology from among an ensemble of plausible chronologies. For example, Li et al. (2011) adjusted age assignments of U/Th-dated fossil Palmyra island corals, observing that the correlation between the nominally-dated modern record, and the first principal component of the North American Drought Atlas (NADA PC1), interpreted as a proxy for ENSO-related hydroclimatic variability, was highly significant over the 1891–1994 time period. They then used this property to align the older sequences, before the onset of the instrumental period, to NADA PC1 within the range of U/Th dating error. To identify an optimal chronology, we similarly choose the realization that maximizes the correlation between

the Palmyra coral record and the NADA PC1 during their periods of overlap (Fig. 2). As in Li et al. (2011), we account for the estimated ±10 yr U/Th dating errors by shifting the final time t_f of the ensemble of plausible chronologies generated with our model across a 21 yr window, so the ensemble size is 20 times larger than in the known t_f case. For convergence purposes, here we chose to use an ensemble size of 10^4 and the same error rates as in the previous section ($\theta_1 = \theta_2 = 0.05$). The selected chronologies (blue curves in the top Fig. 2) improve the correlation coefficient of the individual series to NADA PC1 by a factor of 3 relative to the published chronologies (before dating errors were corrected). Those results are comparable to the ameliorations obtained in Li et al. (2011) after adjustments.

A caveat is of this approach is that it assumes that the NADA PC1 is perfectly known, whereas in reality it is also estimated from layer-counted records. Ideally, one would apply this approach to the NADA record to incorporate uncertainties about this chronology for a full appraisal of uncertainties in this synchronization process and select the chronologies for each that maximize the correlation. However, the large number of cross-dated tree-ring chronologies in NADA (several thousand over this time interval) makes it likely that these uncertainties are negligible compared to those associated with coral dating.

3 Multivariate generalization

Climate proxies are increasingly used as part of large-scale data syntheses (e.g. PAGES2K Consortium, 2013; Marcott et al., 2013; Tierney et al., 2013; Tingley and Huybers, 2013). We thus generalize the previous model to the case of a multivariate network and investigate the effect of age errors on the estimation of spatiotemporal modes of climate variability.

3.1 Model formulation

Let us consider a multivariate banded dataset $\mathbf{B} := \{\boldsymbol{\zeta}, \boldsymbol{\tau}\}$, with $\boldsymbol{\zeta}$ the data matrix ($n \times p$) containing proxy measurements (for instance, $\delta^{18}O$ or Sr/Ca) and $\boldsymbol{\tau}$ the time matrix. Each entry ζ_{ij}, $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2 \ldots, p\}$, corresponds to the ith observation at the jth location, and is assigned a time $\tau_{ij} = t_i^{(j)}$. The previous formulation can be extended by allowing different error parameters θ_{1j} and θ_{2j} for each location j. Then, the matrix of time increments is expressed as:

$$\Delta_{ij} = 1 + P_i^{\theta_{1j}} - \min\left(P_i^{\theta_{2j}}, 1\right) \tag{4}$$

and the chronology matrix verifies:

$$\tau_{ij} = t_f - i - \sum_{k=1}^{i} P_k^{\theta_{1j}} + \sum_{k=1}^{i} \min\left(P_k^{\theta_{2j}}, 1\right) \tag{5}$$

As in Eq. (3), errors are cumulative, and may lead to important age offsets in the earlier portions of the dataset. To see this, consider the situation depicted in Fig. 4, which considers five coherent, monochromatic signals. Age errors on these harmonic signals were simulated according to Eq. (4), using $\theta_{1j} = \theta_{2j} = 0.05$. The figure shows clearly how even small age offsets introduce phase distortion that lowers the coherency between signals, producing an apparent amplitude modulation of the periodic component, whose amplitude appears to grow over time. In Sect. 4 we exploit this property to identify age models based on their ability to correct for this loss of coherency.

3.2 Pseudocoral benchmark

To illustrate the use of the model described above, and to validate our approach, we need a dataset that is free of chronological errors and measurement noise. We thus

consider a "pseudocoral" network of 22 corals, whose location is derived from an existing network described in Table 1, and derived from simulated climate fields. Pseudocoral locations were chosen to mimic the position of publicly available[1] coral δ^{18}O observations with a resolution better than 2 months (Fig. 3).

The pseudocoral observations $\mathbf{C} := \{\boldsymbol{\zeta}, \boldsymbol{\tau}\}$ were simulated according to the forward model of Thompson et al. (2011), a simple representation of the effects of temperature and sea water isotopic variation on the oxygen isotope composition of coral aragonite:

$$\zeta_{ij} = a_1 \mathrm{SST}_{ij} + a_2 \mathrm{SSS}_{ij} + \sigma \cdot \epsilon_{ij} \qquad (6)$$

where a_1 and a_2 describe the dependence of oxygen isotopic equilibrium on the temperature of carbonate formation and regional regression coefficients of SSS onto sea water δ^{18}O respectively, $\epsilon \sim \mathcal{N}(0,1)$ is an uncorrelated standard Gaussian white noise process, and SST and SSS stand for sea surface temperature and sea surface salinity, respectively. The SST and SSS fields were extracted from the CMIP3 "H1" simulation of the Geophysical Fluid Dynamics Laboratory (GFDL) Coupled Model 2.0 (Knutson et al., 2006). The GFDL CM2.0 model is a state-of-the art coupled general circulation model whose performance at simulating tropical Pacific climate is extensively described (Wittenberg et al., 2006; Capotondi et al., 2006; Vecchi et al., 2006; Kug et al., 2010; Xie et al., 2010). Briefly, the model successfully captures the observed annual-mean trade winds and precipitation, sea surface temperature, surface heat fluxes, surface currents, equatorial undercurrent, and subsurface thermal structure. CM2.0 displays a robust El Niño Southern Oscillation (ENSO) with multidecadal fluctuations in amplitude, an irregular period between 2 and 5 yr, and a distribution of SST anomalies that is skewed toward warm events as observed. Although the model is limited by biases common to other coupled GCMs, it provides a virtual laboratory of adequate realism to explore the impact of pseudocoral sampling with chronological errors. The data cover the interval [1860, 2000]; the monthly output was averaged over the DJF months to

---

[1] http://www.ncdc.noaa.gov/paleo/corals.html

sample of the period of most intense ENSO teleconnections. Our pseudocoral coral matrix thus contains $n = 140$ samples at $p = 22$ sites.

In the remainder of the paper, we will use C as a multivariate testing ground for our method. For this study, we limit ourselves to the noiseless case where $\sigma = 0$, though in principle one would have to account for observational noise. Since our interest lies primarily in assessing uncertainties arising from chronological errors, a noise-free matrix is an appropriate first step. We generate an ensemble of age-perturbed realizations of the pseudocoral dataset by disturbing the chronology τ according to the model of Sect. 3.1 using again $\theta_1 = \theta_2 = 0.05$.

3.3 Uncertainty quantification

Inspired by the work of Anchukaitis and Tierney (2012), we first seek to quantify the effect of age uncertainties on the estimation of spatiotemporal modes of variability. Since the CM2.0 model simulates a vigorous ENSO, we seek to answer the following question: were tropical climate variability sampled according to the network of Table 1, how faithfully would ENSO variability be captured in the presence of chronological uncertainties? Figure 3 shows the result of a principal component analysis applied on the error-free network, together with an identical analysis applied to the time-perturbed ensemble. The visualization is analogous to that of Tierney et al. (2013, their Fig. 2).

The unperturbed mode accounts for 25 % of the total variance, and is dominated by the contributions of central and eastern-Pacific pseudocorals. Its principal component (PC) is characterized by a concentration of variance in the interannual band (Fig. 3c, red curves), and its spatial structure (light colored dots with white outline) closely matches the SST pattern obtained by regressing the principal component onto the SST field, meaning that SST dominates SSS in Eq. (6), at least on interannual timescales. The same map presents results from the time-uncertain ensemble, showing that in the presence of time uncertainties the vast majority of positive loadings (dark-colored disks, circled in black) tend to shrink compared to their unperturbed counterparts (light-colored disks, circled in white). The picture is less clear for neg-

ative loadings, but those exert a lesser control on the overall structure. In the age-perturbed ensemble, the amount of variance explained by the ENSO mode ranges between 6 and 10 % (median value: 8 %) down from 25 % for the error-free data. In such cases, the ENSO mode thus ceases to be the leading mode, as most of its en-

⁵ ergy is transferred to a trend mode similar to the one identified by Thompson et al. (2011). In our case, the importance of this mode is mainly an artifact of dating errors. A spectral analysis of the age-perturbed PC (Fig. 3c, blue colors) shows a systematic increase in multidecadal variability (10–50 yr periods) at the expense of interannual variability. The transfer of power from high to low-frequencies may be seen by

¹⁰ computing the spectral slope β, where we fit the model $S(f) \propto f^{\beta}$ to the continuum of each spectrum ($f \leq 1/15$). The spectrum of the unperturbed ENSO PC has a slope $\beta_{\circ} = -0.72$, while the age-perturbed ensemble shows values centered around a median value $\beta_{m} = -0.22$ with an inter-quartile range of $[-0.41, 0]$. The loss of coherency thus has the effect of flattening the spectrum of the ENSO mode. Even when the er-

¹⁵ ror rate is 50 times smaller ($\theta = 0.001$), one still observes that age uncertainties tend to exaggerate multidecadal (20 yr and longer) variations at the expense of interannual variations (Fig. 5). This may explain why Ault et al. (2009) found, using a similar PCA approach, that their coral network exhibited very energetic variability at such scales, with an amplitude that grew back in time. The presence of small age uncertainties

²⁰ could explain part of the multidecadal power enhancement they observed in the late nineteenth century, since fewer instrumental observations were available to provide chronological anchors to coral timeseries at that time than in the twentieth century. Our analysis suggests that the accumulation of age model uncertainty is one possible – non-climatic – explanation for this apparent enhancement of low-frequency variability

²⁵ in an all-coral network, though a more detailed investigation (incorporating knowledge of error rates in each of their records) would be required to evaluate this quantitatively.

4 Coherence-based identification of age models

As in Sect. 2.3, one would like to choose an optimal chronology among the ensemble of competing age models. To do so, we apply an optimization principle based on the idea that all pseudocorals in our network share some information about a com-

⁵ mon signal (ENSO in our pseudocoral benchmark). Indeed, coral networks have been shown to record spatiotemporally coherent climate modes (Evans et al., 1998, 2000, 2002), a property shared by our pseudocoral network benchmark (Fig. 3). Moreover, we showed in Sect. 3.3 that chronological errors lead to a loss of coherency among records; that is, we expect age errors to lower interannual coherency among records

¹⁰ compared to what it might otherwise be. One way to correct for this effect is to pick the age model that maximizes interannual coherency between measurements.

This exercise is however complicated by the large dimensionality of the space of possible solutions. The dimensionality may be estimated in several ways: assuming that age-perturbations follow a Bernoulli process, an $n \times p$ data matrix admits up to 3^{np} age

¹⁵ perturbation matrices. This number quickly becomes astronomical, even for moderate n and p (an NP-hard problem). A more relevant measure, however, is the number of solutions for a given error rate. Given a symmetric error rate $\theta_1 = \theta_2 = \theta$, one expects on average, $\left(\widetilde{\frac{np}{\theta np}} \right)$ (where $\widetilde{x}$ denotes the nearest integer of x) plausible solutions, which induces a serious computational challenge. This is illustrated in Table 2, which charts

²⁰ the number of solutions for various values of np and θ.

In practice, we generate an ensemble of $L = 10^5$ time perturbations $\mathbf{\Delta} = \{\mathbf{\Delta}^{(1)}, \mathbf{\Delta}^{(2)}, \dots, \mathbf{\Delta}^{(L)}\}$ according to Eq. (4), and choose an optimal chronology amongst those realizations. The optimization principle is described next.

4.1 Optimization scheme

²⁵ Because age errors tend to lower coherency amongst records (Fig. 4), the quantity we seek to maximize is the generalized magnitude squared coherency (GMSC,

Chakrabarty et al., 2002; Ramirez et al., 2008, 2010), an extension of the traditional coherency to more than two timeseries. More precisely, we maximize this quantity over the interannual band, where age perturbations have the largest effect on coherency. The method proceeds as described in Algorithm 1.

Algorithm 1: Coherency-based optimization

Input: ζ, L, θ

Output: l^*

1. Generate L age perturbation realizations (2), $\boldsymbol{\Delta} = \{\boldsymbol{\Delta}^{(1)}, \boldsymbol{\Delta}^{(2)}, \ldots, \boldsymbol{\Delta}^{(L)}\}$. Interpolation onto this new grid yields L corrected data matrices $\boldsymbol{\zeta} = \{\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}, \ldots, \boldsymbol{\zeta}^{(L)}\}$.
2. Use SSAM reconstruction to estimate the missing values of each candidate data
3. Compute the GMSC $\hat{\gamma}_l^2$ of each realization $\boldsymbol{\zeta}^{(l)}$.
4. Maximize coherency in the interannual band:

$$l^* = \arg\max_l \left(\int_{\omega_1}^{\omega_2} \hat{\gamma}_l^2(\omega)\, d\omega \right) \tag{7}$$

with $\omega_1 = 2\pi/(8\,\mathrm{yr}), \omega_2 = 2\pi/(2\,\mathrm{yr})$

4.2 Method validation

To check that the above principle can identify the correct age model, we employ the following strategy. A reference age-perturbation process is used to alter the chronology of the error-free pseudocoral dataset described in Sect. 3.3. We then generate an ensemble of age-corrections (Sect. 3.1), for the reference perturbed dataset $\mathbf{C}^* := \{\boldsymbol{\zeta}^*, \boldsymbol{\tau}^*\}$ yielding an ensemble of candidate age-corrected datasets, and apply the optimization

6093

scheme described above. We then verify the validity of our model by comparing the retained chronology to the true one.

For a given θ, the verification algorithm is explained in Algorithm 2.

Algorithm 2: Age model identification

Input: ζ, L, θ

Output: Confidence Interval, Coverage probability, l^*

1. Generate a reference age perturbation $\mathbf{P}^*$ and corresponding chronology $\boldsymbol{\tau}^*$ with $\theta_1 = \theta_2 = \theta$
2. Interpolate data $\boldsymbol{\zeta}$ onto $\boldsymbol{\tau}^*$ using SSAM, yielding $\boldsymbol{\zeta}^*$
3. Generate L correction matrices (equation 4) $\boldsymbol{\Delta} = \{\boldsymbol{\Delta}^{(l)}; l = 1, 2, \ldots, L\}$ and corresponding corrected candidate data $\boldsymbol{\zeta_c} = \{\boldsymbol{\zeta}_c^{(l)}; l = 1, 2, \ldots, L\}$ from reference data $\boldsymbol{\zeta}^*$
4. Compute 95% confidence intervals for $\boldsymbol{\zeta_c}$
5. Compute the coverage probability of $\boldsymbol{\zeta_c}$
6. Find l^* from Alg. 1

The results (Fig. 6) show the rapid loss of coherency over the interannual band as more age errors (increasing θ parameter) are introduced in the time series. Indeed, the $\theta = 0.001$ case brings the vast majority of the time-uncertain ensemble spectra under the coherence spectrum of the original time series, and the $\theta = 0.05$ case suppresses nearly all the coherency for periods shorter than 4 yr, and greatly subdues the original peak around 5 yr.

For $\theta = 0.05$, the optimization principle failed to identify the correct age model, as no single realization of the corrected ensemble presented high GMSC values in the interannual band. The issue stems from the very high dimensionality of the space of plausible age-corrections (Table 2), which results in the inability of our sampled ensembles to span the solution space effectively. We attempted to use a Latin Hypercube

6094

sampling scheme in an effort to cover the solution space more effectively, however we did not notice any major improvement compared to the Monte-Carlo ensemble results. More work is required here to find efficient sampling techniques.

One way to evaluate the potential of this method to accurately portray uncertainties is to compute the actual coverage rates (Guttman, 1970) of confidence intervals (CIs) derived from our ensemble. To do so, we evaluate the probabilities that the ensemble CIs cover the original dataset, and compare those to their nominal value of 95 (Table 3). As θ increases, so do time uncertainties, thereby widening the confidence interval of the ensemble. The coverage probability remains high for all θ, since the error-free data lies within the corrected ensemble nearly 98 % of the time. Our CIs are thus slightly conservative.

While the difficulty in solving this NP-hard optimization problem gives little hope of finding the correct age model even for the smallest θ (note from Table 2 that the expected number of plausible chronologies to correct 1 % age errors is on the order of 10^{15}), we here show that the coherence-based optimization principle can indeed bring the age model closer to the correct one.

To see this quantitatively, one first needs to define a distance over age model space. The Hamming distance $d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y})$ (Hamming, 1950) between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$ provides such a metric. $d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y})$ is defined as the proportion of corresponding coefficients positions in which they differ, i.e.

$$d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n} \sum_{i=1}^{n} d_i$$

where

$$d_i = \begin{cases} 1 & \text{if } x_i = y_i \\ 0 & \text{else} \end{cases} \tag{8}$$

Because departures from unity in $\boldsymbol{\Delta}$ are of prime interest, this metric is a good one to differentiate time-shifted records from one another. We generalize the Hamming

6095

distance to our matrix case simply by computing and adding together the distances at each location j, then dividing by p.

In the following experiment, we sample ensembles of chronologies in a restricted neighborhood around a reference age-perturbed dataset $\mathbf{C}^*$. In this context, for each time perturbation in $\mathbf{C}^*$, we sample an age-correction within a neighborhood of size $2 \cdot N_v + 1$, centered around a known miscounted band.

Four age-corrected ensembles are generated using a neighborhood size expanding from $N_v = 1$ to $N_v = 10$ annual bands on either side of the true one. For each sampled realization, we evaluate the Hamming distance to the error-free chronology and compute the global coherence spectrum. For each ensemble, we construct its corresponding Hamming distance distribution and extract the realization that maximizes the GMSC according to algorithm 1. In Fig. 7, the results for each ensemble is color-coded so that one can relate the coherence spectra to the Hamming distance distributions. The green color refers to the reference perturbed dataset with an error rate $\theta = 0.05$ and the red is reserved for the original time series. The dark blue, blue, pink and light brown show the results for $N_v = 1, 4, 7$ and 10, respectively. We observe an increase in the mean of the distance distribution as the size of the neighborhood N_v expands, yet for all the realizations, the distance to the error-free chronology was reduced relative to the distance between the reference and the true chronologies. This experiment shows that when the age-model ensemble includes the true solution, the optimization principle tends to effectively correct the errors, in the sense that it reduces the Hamming distance between the original and corrected chronologies. One the other hand, the realization that maximizes the GMSC for each neighborhood ensemble does not consistently minimize the Hamming distance (see the vertical bars of the ensemble color in the bottom Fig. 7b), however, it, on average, appears on the lower half of each Hamming distance distribution. The black curve in the coherence spectra figure and the line of the same color in the Hamming distance plot refer to the realization that achieves the maximum GMSC value, selected from an unrestricted ensemble such as one generated according to Eq. (5). The sampling issue associated with the extremely

6096

high dimensionality of the solution space can be readily observed here, noticing that the realization maximizing the GMSC barely surpasses the reference coherence spectrum in addition to its distance to the original chronology exceeding the distance between the reference and true data. The issue may be alleviated if there is some prior information about the timing of age errors, for instance an observed mismatch between how two series portray an event – like a historic volcanic eruption – that should be synchronous.

5   Discussion

In this article, we put forth a probabilistic model for layer-counted chronologies. The model was designed for univariate as well as multivariate datasets, and is flexible enough to accommodate location-specific and asymmetric error rates. We used this model to generate ensembles of time-uncertain datasets, which we analyzed using spectral methods and principal component analysis.

In the univariate case, we showed that age-perturbations tend to blur variability between interannual and multidecadal scales. In the multivariate case, we showed that even small age offsets between records may significantly lower coherency between records, producing an apparent amplitude modulation of high-frequency signals. These short-term phase interferences thus led to long-term effects, exaggerating low-frequency variability at the expense of interannual signals. This suggests that at least part of the reason for the observed nineteenth century increase in the amplitude of low-frequency variability (Ault et al., 2009) could be the presence of age errors.

We also attempted to correct for age errors using an optimization principle based on the global coherence of a multivariate record. The large dimensionality of the age model space, however, made the random sampling scheme an inefficient method of exploration. There is no easy way around this problem. Nonetheless, we showed that our optimization strategy was able to correct dating errors when sufficient restrictions were placed on the solution space, provided the mutual information was sufficiently strong.

There are important limitations to our study. First and foremost, our framework assumes that the error rates are known. Ideally, one would estimate the parameters (θ_1, θ_2) from the data, but this is only possible in the presence of independent information (e.g. replicated records, or absolute chronostratigraphic markers like historically-dated events DeLong et al., 2007, 2013). We also restricted our analysis to symmetric error rates, which is an unlikely scenario according to DeLong et al. (2013) who emphasized that double counting is typically more common than missing years. Considering that the events perturbing the chronologies are hidden states of the process and are not being directly measured, a maximum likelihood principle would have to be employed to evaluate the error rate, which substantially complicates the problem of designing a likelihood function. However, when the common signal driving the record is strong enough, a coherency-based Bayesian approach could potentially learn from the data. In this case, one would relax the assumption of known rate parameters (θ_1, θ_2) in favor of prior distributions on those parameters, which would incorporate as much scientific information as possible. In this regard, one should note that most geochronologies are developed with the purpose of producing the single best representation of past changes. We suggest that there would be merit in also providing raw chronologies so that the error rate may be estimated via comparing it with other, absolutely dated records (for instance, instrumental observations at a nearby site, if available). In other words, we hope to encourage field scientists to report these error rates rather than exclusively seeking to reduce them, as this would greatly aid the quantification of uncertainties associated with each record.

Second, we assumed that time perturbations follow a homogeneous Poisson process, with a uniform rate for all pseudocorals considered and assumed that they are symmetric, which is unlikely. In nature, each record will display a different rate, and in situations of uneven dating conditions (e.g. annual bands being more or less visible in a given time interval), an inhomogeneous Poisson process would constitute a more suitable model within a single record. The error rate could be informed by local expertise and change with the location of the data as well as with time and account for known

climatic events that could have affected the coral or breaks, boring and growth structures that are present in the core. Assessments of where and when corals might have over or under-grown impose restrictions on the age model space, which would greatly improve the ability of our optimization principle to identify the correct chronology.

Third, we assumed that age perturbations taking place at different locations are independent. Nonetheless, one could easily conceive of a scenario in which a regional phenomenon would simultaneously affect the growth of all corals located in one area. The model would therefore have to include regional random time perturbations, which could be done by specifying parameters conditional on the climate.

Fourth, we assumed that all records considered had annual resolution, while in reality it could range from a month to several years. Adapting the method to a network of variable resolution would require the conditioning of the rate parameter θ on the resolution of the record's chronology, as for instance, monthly data tends to present a lower error rate than yearly data, but would add subannual errors from selecting incorrect tie points in the seasonal cycle particularly in regions where the seasonal cycle and annual density banding are weak (e.g., key ENSO sensitive regions of the central Pacific).

Fifth, we assumed that the latest time t_f is known exactly (except in Sect. 2.3), which is only true of records that grew continuously until that point (e.g. living corals, living trees). We could refine our fossil corals framework with a mixed approach, where the age of tie points would be allowed to vary according to their distribution (e.g. a Gaussian), and the layer-count model would be extended on either side.

Finally, our optimization scheme was shown to be effective only in extremely restricted situations (error-free measurements of δ^{18}O, age model ensemble spanning the true age model). It remains to be seen whether these conditions could ever be met in practice. One potential problem is the fact that the "clock" used to synchronize the records does not beat periodically: ENSO is only a quasi-periodic phenomenon, even in the GFDL CM2.0 model. Perhaps a more regular cycle (annual, semi-annual) would provide stronger synchronization constraints (but it would limit the analysis to records

that have seasonal or better resolution). Alternatively, one could restrict the optimization process to small, regional ensembles of proxies that share a stronger common signal, much like dendrochronological approaches.

With these caveats in mind, several points deserve mention. First, the simple example of Fig. 4 demonstrates how miscounting errors may produce an apparent amplitude modulation of otherwise stationary signals, producing spurious trends in the average derived from them. Because climate reconstruction methods rely in some fashion on a weighted average of observations assumed to be contemporaneous, this may partially explain why many ENSO reconstructions (e.g. Mann et al., 2000; Wilson et al., 2010; McGregor et al., 2010; Li et al., 2011; Emile-Geay et al., 2013) display increasing trends in ENSO variance over time (McGregor et al., 2013). As pointed out in the latter study, computing ENSO variance on individual records prior to averaging them together is more robust to chronological errors. In the meantime, interpreting trends in variance obtained from the average of age-uncertain records as reflecting a response to external forcing may be perilous. At the very least, a probabilistic model such as the one proposed here can provide a null hypothesis against which to assess such trends.

More generally, a model such as this one may be used to evaluate the plausible impact of dating uncertainties on all manner of inferences derived from time-uncertain records. Just as proxy system models are now becoming a standard tool to interpret environmental controls in paleoclimate proxies, chronological models should become an integral component of proxy system models for age-uncertain paleodata, so that they may be used to improve interpretation, reconstruction, uncertainty estimation, and data/model comparison studies (Evans et al., 2013). Because chronological errors might themselves depend on climatic events (see introduction), a complete proxy system model should in fact allow environmental controls to influence chronological information.

We hope this study stimulates a discussion on how to best represent chronological uncertainty in layer-counted proxy archives, and leads to a better characterization of these uncertainties.

Appendix A

Stochastic model derivation

The stochastic model used to characterize the age uncertainties may be derived as follows: let us denote by $\{N_i : i = 1, 2, \cdots, n\}$, n independent and identically distributed random variables on the probability space $\{\Omega, \Sigma, \mathrm{Pr}\}$ representing the miscounting events on observed layers. Each N_i is a counting process on $[t_i, t_{i+1})$ taking values in the sample set Ω of non-negative integers and $N_i = 0$ if the layer was counted accurately. Let the stochastic process $\{N(t) : t_1 \leq t \leq t_n\}$ equal the number of age perturbations in our data in the time interval $[t_1, t_n]$, then $N(t_i) = \sum_{k=1}^{i} N_k$ is also a counting process. Because of the earlier assumption on the most recent observation being known exactly, we can write that $N(t_1) = 0$. We also parameterize the miscounting error rate by θ such that $E(N_i) = \theta$ and we can prescribe our simulations to accumulate on average θn perturbations. Recognizing that the chances of miscounting more than two years per band are slight, one may write:

$$\mathrm{Pr}\{N(t + h) - N(t) = 1\} = \theta h + o(h) \tag{A1}$$

$$\mathrm{Pr}\{N(t + h) - N(t) \geq 2\} = o(h) \tag{A2}$$

where $f(h) = o(h)$ is a function such that $\frac{f(h)}{h}$ goes to zero with h. One recognizes $N(t)$ as a Poisson process with rate parameter θ (e.g. Suhov and Kelbert, 2008, theorem 2.3.2b). When the level of confidence in a particular chronology is such that $\mathrm{Pr}(N_i) > 1$ is negligible, a Bernoulli process would be a suitable approximation for $N(t)$ with $\mathrm{Pr}(N_i = 1) = \theta$, and $N(t_i)$ would follow a Binomial distribution $\mathcal{B}(i, \theta)$. Age perturbations when counting layers fall in one of two categories: one can either miss (type 1) or double-count a band (type 2), so we classify perturbation events into those two types. Let $\{P_i^{\theta_1} : i = 1, 2, \cdots, n\}$ count the number of missing layers between layers $i - 1$ and i and $\{P_i^{\theta_2} : i = 1, 2, \cdots, n\}$ count the number of duplications of layer i with rate parameter θ_1 and θ_2 respectively, then, if we assume that the two types of events are independent,

we can write that $N_i = P_i^{\theta_1} + P_i^{\theta_2}$ and $\theta = \theta_1 + \theta_2$. The assumption of independence simplifies the problem's treatment, though it is obviously an idealization. In practice, the probability of a layer being counted as part of both of two categories is generally so low that it does not materially affect the results.

A stochastic parameterization of the increment matrix may thus be expressed as follows:

$$\forall i \in [1, n-1], \quad \Delta_i = 1 + P_i^{\theta_1} - \min\left(P_i^{\theta_2}, 1\right) \tag{A3}$$

where the perturbations $\{P_i^{\theta_k} : k = 1, 2\}$ are random variables representing Poisson increments in $[t_i, t_{i+1})$ with rate $\theta_k > 0$ or the outcomes of a Bernoulli process with trial probability θ_k. In both instances, one may choose parameters so that the variables are zero most of time, except for a relatively small fraction (say 5 %) when they are positive, so $\Delta_i > 1$ when $P_i^{\theta_1} > 0$ (i.e. a year is missing) or $\Delta_i < 1$ when $P_i^{\theta_2} > 0$ (i.e. a band was doubled). It follows that the chronology verifies:

$$t_i = t_f - i - \sum_{k=1}^{i} P_k^{\theta_1} + \sum_{k=1}^{i} \min\left(P_k^{\theta_2}, 1\right) \tag{A4}$$

In the Poisson case, if we denote by $N_1(t)$ and $N_2(t)$ the number of events $\sum_k P_k^{\theta_1}$ and $\sum_k P_k^{\theta_2}$ until time t respectively, then the difference between the two Poisson processes $N_1(t) - N_2(t)$ is known to follow a Skellam distribution (Skellam, 1946) with parameters $\lambda_1 = \theta_1 t$ and $\lambda_2 = \theta_2 t$, which is given by:

$$\mathrm{Pr}\{N_1(t) - N_2(t) = k\} = e^{-(\lambda_1 + \lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{k/2} \mathbf{I}_{|k|}(2\sqrt{\lambda_1 \lambda_2}) \tag{A5}$$

where $\mathbf{I}_k(x)$ is a modified Bessel function of the first kind (e.g. Abramowitz and Stegun, 1965). The distribution is plotted in Fig. A1, and we here summarize its main properties:

1. its mean is equal to $\lambda_1 - \lambda_2$, its variance to $\lambda_1 + \lambda_2$.

2. for $\lambda_1 = \lambda_2$, the distribution is symmetric, centered at zero.

3. for $\lambda_1 = \lambda_2$ large enough, it is well approximated by a normal distribution $\mathcal{N}(0, \lambda_1 + \lambda_2)$, thus the difference between the Poisson model and the binomial model (which also converges to a normal distribution) becomes negligible.

References

Abram, N. J., Gagan, M. K., Cole, J. E., Hantoro, W. S., and Mudelsee, M.: Recent intensification of tropical climate variability in the Indian Ocean, Nat. Geosci., 1, 849–853, doi:10.1038/ngeo357, 2008. 6111

Abramowitz, M. and Stegun, I. A.: Handbook of Mathematical Functions, National Bureau of Standards, Applied Math #55, Dover Publications, New York, 1965. 6102

Alley, R. B., Shuman, C. A., Meese, D. A., Gow, A. J., Taylor, K. C., Cuffey, K. M., Fitzpatrick, J. J., Grootes, P. M., Zielinski, G. A., Ram, M., Spinelli, G., and Elder, B.: Visual-stratigraphic dating of the GISP2 ice core: basis, reproducibility, and application, J. Geophys. Res.-Atmos., 102, 26367–26381, doi:10.1029/96JC03837, 1997. 6080, 6081

Anchukaitis, K. J. and Tierney, J. E.: Identifying coherent spatiotemporal modes in time-uncertain proxy paleoclimate records, Clim. Dynam., 41, 1–16, doi:10.1007/s00382-012-1483-0, 2012. 6081, 6090

Asami, R., Yamada, T., Iryu, Y., Quinn, T. M., Meyer, C. P., and Pauley, G.: 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, 2005. 6111

Ault, T. R., Cole, J. E., Evans, M. N., Barnett, H., Abram, N. J., Tudhope, A. W., and Linsley, B. K.: Intensified decadal variability in tropical climate during the late 19th century, Geophys. Res. Lett., 36, L08602, doi:10.1029/2008GL036924, 2009. 6091, 6097

Bender, M. L.: Orbital tuning chronology for the Vostok climate record supported by trapped gas composition, Earth Planet. Sc. Lett., 204, 275–289, doi:10.1016/S0012-821X(02)00980-9, 2002. 6080

Blaauw, M.: Methods and code for "classical" age-modelling of radiocarbon sequences, Quat. Geochronol., 5, 512–518, doi:10.1016/j.quageo.2010.01.002, 2010. 6081

Blaauw, M.: Out of tune: the dangers of aligning proxy archives, Quaternary Sci. Rev., 36, 38–49, doi:10.1016/j.quascirev.2010.11.012, 2012. 6081

Blaauw, M. and Christen, J. A.: Flexible paleoclimate age-depth models using an autoregressive gamma process, Bayesian Anal., 6, 457–474, 2011. 6081

Boiseau, M., Juillet-Leclerc, A., Yiou, P., Salvat, B., Isdale, P., and Guillaume, M.: Atmospheric and oceanic evidences of El Niño-Southern Oscillation events in the south central Pacific Ocean from coral stable isotopic records over the last 137 yr, Paleoceanography, 13, 671–685, doi:10.1029/98PA02502, 1998. 6111

Broomhead, D. and King, G. P.: Extracting qualitative dynamics from experimental data, Physica D, 20, 217–236, doi:10.1016/0167-2789(86)90031-X, 1986. 6085

Capotondi, A., Wittenberg, A., and Masina, S.: Spatial and temporal structure of tropical Pacific interannual variability in 20th century coupled simulations, Ocean Model., 15, 274–298, 2006. 6089

Chakrabarty, K., Sitharama Iyengar, S., Qi, H., and Cho, E.: Grid coverage for surveillance and target location in distributed sensor networks, IEEE T. Comput., 51, 1448–1453, 2002. 6093

Charles, C. D., Hunter, D. E., and Fairbanks, R. G.: 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, 1997. 6111

Charles, C. D., Cobb, K., Moore, M. D., and Fairbanks, R. G.: Monsoon-tropical ocean interaction in a network of coral records spanning the 20th century, Mar. Geol., 201, 207–222, doi:10.1016/S0025-3227(03)00217-2, 2003. 6111

Chave, A. D., Thomson, D. J., and Ander, M. E.: On the robust estimation of power spectra, coherences, and transfer functions, J. Geophys. Res., 92, 633–648, doi:10.1029/JB092iB01p00633, 1987. 6086

Cobb, K. M., Charles, C. D., Cheng, H., and Edwards, R. L.: El Niño/Southern Oscillation and tropical Pacific climate during the last millennium, Nature, 424, 271–276, 2003. 6080, 6086, 6114, 6116

Cobb, K. M., Charles, C. D., and Hunter, D. E.: A central tropical Pacific coral demonstrates Pacific, Indian, and Atlantic decadal climate connections, Geophys. Res. Lett., 28, 2209–2212, doi:10.1029/2001GL012919, 2001. 6111

Cook, E. R. and Kairiukstis, L. A.: Methods of Dendrochronology: Applications in the Environmental Sciences, Kluwer Academic Publishers, Doredrecht, 1990. 6080

Damassa, T. D., Cole, J. E., Barnett, H. R., Ault, T. R., and McClanahan, T. R.: Enhanced multidecadal climate variability in the seventeenth century from coral isotope records in the western Indian Ocean, Paleoceanography, 21, PA2016, doi:10.1029/2005PA001217, 2006. 6111

DeLong, K. L., Quinn, T. M., and Taylor, F. W.: Reconstructing twentieth-century sea surface temperature variability in the southwest Pacific: a replication study using multiple coral Sr/Ca records from New Caledonia, Paleoceanography, 22, PA4212, doi:10.1029/2007PA001444, 2007. 6082, 6098

DeLong, K. L., Flannery, J. A., Maupin, C. R., Poore, R. Z., and Quinn, T. M.: A coral Sr/Ca calibration and replication study of two massive corals from the Gulf of Mexico, Palaeogeogr. Palaeocl., 307, 117–128, 2011. 6082

DeLong, K. L., Quinn, T. M., Taylor, F. W., Shen, C.-C., and Lin, K.: Improving coral-base paleoclimate reconstructions by replicating 350 yr of coral Sr/Ca variations, Palaeogeogr. Palaeocl., 373, 6–24, 2013. 6082, 6083, 6085, 6098

Douglass, A. E.: Crossdating in dendrochronology, J. Forestry, 39, 825–831, 1941. 6080

Emile-Geay, J. and Eshleman, J. A.: Toward a semantic web of paleoclimatology, Geochem. Geophy. Geosy., 14, 457–469, doi:10.1002/ggge.20067, 2013. 6111

Emile-Geay, J., Cobb, K., Mann, M., and Wittenberg, A. T.: Estimating central equatorial Pacific SST variability over the past millennium, Part 2: Reconstructions and implications, J. Climate, 26, 2329–2352, doi:10.1175/JCLI-D-11-00511.1, 2013. 6081, 6100

Evans, M. N., Kaplan, A., and Cane, M. A.: Optimal sites for coral-based reconstruction of global sea surface temperature, Paleoceanography, 13, 502–516, doi:10.1029/98PA02132, 1998. 6092

6105

Evans, M. N., Kaplan, A., and Cane, M. A.: Intercomparison of coral oxygen isotope data and historical sea surface temperature (SST): potential for coral-based SST field reconstructions, Paleoceanography, 15, 551–562, 2000. 6092

Evans, M. N., Kaplan, A., and Cane, M. A.: Pacific sea surface temperature field reconstruction from coral δ^{18}O data using reduced space objective analysis, Paleoceanography, 17, 7–1, doi:10.1029/2000PA000590, 2002. 6092

Evans, M., Tolwinski-Ward, S., Thompson, D., and Anchukaitis, K.: Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Sci. Rev., 76, 16–28, 2013. 6100

Fairbanks, R. G. and Dodge, R. E.: Annual periodicity of the ^{18}O/^{16}O and ^{13}C/^{12}C ratios in the coral *Montastrea annularis*, Geochim. Cosmochim. Acta, 43, 1009–1020, 1979. 6080

Felis, T., Pätzold, J., Loya, Y., Fine, M., Nawar, A. H., and Wefer, G.: 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, doi:10.1029/1999PA000477, 2000. 6111

Fraedrich, K.: Estimating the dimensions of weather and climate attractors, J. Atmos. Sci., 43, 419–432, doi:10.1175/1520-0469(1986)043<0419:ETDOWA>2.0.CO;2, 1986. 6085

Fritts, H. C.: Tree Rings and Climate, Academic Press, 567 pp., London, 1976. 6080

Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S., and Jouzel, J.: Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores, Nature, 366, 552–554, doi:10.1038/366552a0, 1993. 6080

Guttman, I.: Statistical Tolerance Regions: Classical and Bayesian, vol. 26, Griffin's statistical monographs series, Publisher: Hafner Pub. Co. the University of Michigan, 1970. 6095

Haam, E. and Huybers, P.: A test for the presence of covariance between time-uncertain series of data with application to the Dongge Cave speleothem and atmospheric radiocarbon records, Paleoceanography, 25, PA2209, doi:10.1029/2008PA001713, 2010. 6081

Hamming, R. W.: Error detecting and error correcting codes, Bell Syst. Tech. J., 29, 147–160, 1950. 6095

Hays, J. D., Imbrie, J., and Shackleton, N. J.: Variations in the Earth's orbit: pacemaker of the ice ages, Science, 194, 1121–1132, 1976. 6080

Huybers, P. and Wunsch, C.: A depth-derived Pleistocene age model: Uncertainty estimates, sedimentation variability, and nonlinear climate change, Paleoceanography, 19, PA1028, doi:10.1029/2002PA000857, 2004. 6081

6106

Isdale, P. J., Stewart, B. J., Tickle, K. S., and Lough, J. M.: Palaeohydrological variation in a tropical river catchment: a reconstruction using fluorescent bands in corals of the Great Barrier Reef, Australia, The Holocene, 8, 1–8, doi:10.1191/095968398670905088, 1998. 6085, 6114

Knutson, T. R., Delworth, T. L., Dixon, K. W., Held, I. M., Lu, J., Ramaswamy, V., Schwarzkopf, M. D., Stenchikov, G., and Stouffer, R. J.: Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models, J. Climate, 19, 1624–1651, doi:10.1175/JCLI3709.1, 2006. 6089

Kug, J.-S., Choi, J., An, S.-I., Jin, F.-F., and Wittenberg, A. T.: Warm pool and cold tongue El Niño events as simulated by the GFDL CM2.1 coupled GCM, J. Climate, 23, 1226–1239, doi:10.1175/2009JCLI3293.1, 2010. 6089

Kuhnert, H., Pätzold, J., Hatcher, B., Wyrwoll, K. H., Eisenhauer, A., Collins, L. B., Zhu, Z. R., and Wefer, G.: A 200 yr coral stable oxygen isotope record from a high-latitude reef off Western Australia, Coral Reefs, 18, 1–12, doi:10.1007/s003380050147, 1999. 6111

Kuhnert, H., Pätzold, J., Wyrwoll, K.-H., and Wefer, G.: Monitoring climate variability over the past 116 yr in coral oxygen isotopes from Ningaloo Reef, Western Australia, Int. J. Earth Sci., 88, 725–732, doi:10.1007/s005310050300, 2000. 6111

Li, J., Xie, S.-P., Cook, E. R., Huang, G., D'Arrigo, R., Liu, F., Ma, J., and Zheng, X.-T.: Interdecadal modulation of El Niño amplitude during the past millennium, Nat. Clim. Change, 1, 114–118, 2011. 6086, 6087, 6100

Linsley, B. K., Dunbar, R. B., Wellington, G. M., and Mucciarone, D. A.: A coral-based reconstruction of intertropical convergence zone variability over Central America since 1707, J. Geophys. Res., 99, 9977–9994, doi:10.1029/94JC00360, 1994. 6111

Linsley, B. K., Ren, L., Dunbar, R. B., and Howe, S. S.: El Niño Southern Oscillation (ENSO) and decadal-scale climate variability at 10° N in the eastern Pacific from 1893 to 1994: a coral-based reconstruction from Clipperton Atoll, Paleoceanography, 15, 322–335, doi:10.1029/1999PA000428, 2000. 6111

Linsley, B. K., Kaplan, A., Gouriou, Y., Salinger, J., deMenocal, P. B., Wellington, G. M., and Howe, S. S.: Tracking the extent of the South Pacific Convergence Zone since the early 1600s, Geochem. Geophy. Geosy., 7, Q05003, 2006. 6111

Linsley, B. K., Zhang, P., Kaplan, A., Howe, S. S., and Wellington, G. M.: Interdecadal-decadal climate variability from multicoral oxygen isotope records in the South Pacific Convergence Zone region since 1650 AD, Paleoceanography, 23, PA2219, doi:10.1029/2007PA001539, 2008. 6111

6107

Lisiecki, L. E. and Raymo, M. E.: A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18}O records, Paleoceanography, 20, PA1003, doi:10.1029/2004PA001071, 2005. 6080

Mann, M., Gilleb, E., Bradley, R., Hughes, M., Overpeck, J., Keimigc, F., and Gross, W.: Global temperature patterns in past centuries: an interactive presentation, Earth Interact., 4, 1–29, 2000. 6100

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A reconstruction of regional and global temperature for the past 11 300 yr, Science, 339, 1198–1201, doi:10.1126/science.1228026, 2013. 6087

Martinson, D. G., Menke, W., and Stoffa, P.: An inverse approach to signal correlation, J. Geophys. Res., 87, 4807–4818, doi:10.1029/JB087iB06p04807, 1982. 6080

McGregor, S., Timmermann, A., and Timm, O.: A unified proxy for ENSO and PDO variability since 1650, Clim. Past, 6, 1–17, doi:10.5194/cp-6-1-2010, 2010. 6100

McGregor, S., Timmermann, A., England, M. H., Elison Timm, O., and Wittenberg, A. T.: Inferred changes in El Niño–Southern Oscillation variance over the past six centuries, Clim. Past, 9, 2269–2284, doi:10.5194/cp-9-2269-2013, 2013. 6100

Mudelsee, M., Scholz, D., Röthlisberger, R., Fleitmann, D., Mangini, A., and Wolff, E. W.: Climate spectrum estimation in the presence of timescale errors, Nonlinear Proc. Geoph., 16, 43–56, doi:10.5194/npg-16-43-2009, 2009. 6085

PAGES2K Consortium: Continental-scale temperature variability during the past two millennia, Nat. Geosci, 6, 339–346, doi:10.1038/ngeo1797, 2013. 6087

Pfeiffer, M., Timm, O., Dullo, W.-C., and Podlech, S.: Oceanic forcing of interannual and multidecadal climate variability in the southwestern Indian Ocean: evidence from a 160 yr coral isotopic record (La Réunion, 55° E, 21° S), Paleoceanography, 19, PA4006, doi:10.1029/2003PA000964, 2004. 6111

Quinn, T. M., Crowley, T. J., Taylor, F. W., Henin, C., Joannot, P., and Join, Y.: A multicentury stable isotope record from a New Caledonia coral: interannual and decadal sea surface temperature variability in the southwest Pacific since 1657 AD, Paleoceanography, 13, 412–426, doi:10.1029/98PA00401, 1998. 6111

Quinn, T. M., Taylor, F. W., and Crowley, T. J.: Coral-based climate variability in the Western Pacific Warm Pool since 1867, J. Geophys. Res.-Oceans, 111, C11006, doi:10.1029/2005JC003243, 2006. 6111

Ramirez, D., Via, J., and Santamaria, I.: A generalization of the magnitude squared coherence spectrum for more than two signals: definition, properties and estimation, in: IEEE Interna-

6108

tional Conference on Acoustics, Speech and Signal Processing, 2008 (ICASSP 2008), IEEE, 3769–3772, doi:10.1109/ICASSP.2008.4518473, 2008. 6093

Ramirez, D., Via, J., Santamaria, I., and Scharf, L. L.: Detection of spatially correlated Gaussian time series, IEEE T. Signal Proces., 58, 5006–5015, 2010. 6093

Rhines, A. and Huybers, P.: Estimation of spectral power laws in time uncertain series of data with application to the Greenland Ice Sheet Project 2 δ^{18}O record, J. Geophys. Res.-Atmos., 116, D01103, doi:10.1029/2010JD014764, 2011. 6081

Schoellhamer, D. H.: Singular spectrum analysis for time series with missing data, Geophys. Res. Lett., 28, 3187–3190, doi:10.1029/2000GL012698, 2001. 6085

Seimon, A.: Improving climatic signal representation in tropical ice cores: a case study from the Quelccaya Ice Cap, Peru, Geophys. Res. Lett., 30, 1772, doi:10.1029/2003GL017191, 2003. 6081

Skellam, J. G.: The frequency distribution of the difference between two Poisson variates belonging to different populations, J. R. Stat. Soc., 109, p. 296, 1946. 6102

St. George, S., Ault, T. R., and Torbenson, M. C. A.: The rarity of absent growth rings in Northern Hemisphere forests outside the American Southwest, Geophys. Res. Lett., 40, 3727–3731, doi:10.1002/grl.50743, 2013. 6080

Stokes, M. A. and Smiley, T. L.: An Introduction to Tree-Ring Dating, The University of Chicago Press, Chicago, Illinois, 1996. 6080

Suhov, Y. and Kelbert, M.: Probability and Statistics by Example: Markov Chains: a Primer in Random Processes and Their Applications, Cambridge University Press, New York, 2008. 6101

Thompson, D. M., Ault, T. R., Evans, M. N., Cole, J. E., and Emile-Geay, J.: Comparison of observed and simulated tropical climate trends using a forward model of coral of δ^{18}O, Geophys. Res. Lett., 38, L14706, doi:10.1029/2011GL048224, 2011. 6089, 6091

Thompson, L. G., Mosley-Thompson, E., Bolzan, J. F., and Koci, B. R.: A 1500 yr record of tropical precipitation in ice cores from the Quelccaya Ice Cap, Peru, Science, 229, 971–973, doi:10.1126/science.229.4717.971, 1985. 6081

Thomson, D. J.: Spectrum estimation and harmonic analysis, Proc. IEEE, 70, 1055–1096, 1982. 6085

Tierney, J. E., Smerdon, J. E., Anchukaitis, K. J., and Seager, R.: Multidecadal variability in East African hydroclimate controlled by the Indian Ocean, Nature, 493, 389–392, doi:10.1038/nature11785, 2013. 6087, 6090

6109

Tingley, M. P. and Huybers, P.: Recent temperature extremes at high northern latitudes unprecedented in the past 600 yr, Nature, 496, 201–205, doi:10.1038/nature11969, 2013. 6087

Tudhope, A. W., Chilcott, C. P., McCulloch, M. T., Cook, E. R., Chappell, J., Ellam, R. M., Lea, D. W., Lough, J. M., and Shimmield, G. B.: Variability in the El Niño–Southern Oscillation through a glacial–interglacial cycle, Science, 291, 1511–1517, doi:10.1126/science.1057969, 2001. 6111

Urban, F. E., Cole, J. E., and Overpeck, J. T.: Influence of mean climate change on climate variability from a 155 yr tropical Pacific coral record, Nature, 407, 989–993, 2000. 6111

Vautard, R. and Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Physica D, 35, 395–424, 1989. 6085

Vecchi, G. A., Soden, B. J., Wittenberg, A. T., Held, I. M., Leetmaa, A., and Harrison, M. J.: Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing, Nature, 441, 73–76, 2006. 6089

Wang, Y. J., Cheng, H., Edwards, R. L., An, Z. S., Wu, J. Y., Shen, C.-C., and Dorale, J. A.: A high-resolution absolute-dated late Pleistocene Monsoon record from Hulu Cave, China, Science, 294, 2345–2348, doi:10.1126/science.1064618, 2001. 6080

Wigley, T. M. L., Jones, P. D., and Briffa, K. R.: Cross-dating methods in dendrochronology, J. Archaeol. Sci., 14, 51–64, 1987. 6080

Wilson, R., Cook, E., D'Arrigo, R., Riedwyl, N., Evans, M. N., Tudhope, A., and Allan, R.: Reconstructing ENSO: the influence of method, proxy data, climate forcing and teleconnections, J. Quaternary Sci., 25, 62–78, doi:10.1002/jqs.1297, 2010. 6100

Wittenberg, A. T., Rosati, A., Lau, N.-C., and Ploshay, J. J.: GFDL's CM2 global coupled climate models, Part 3: Tropical Pacific climate and ENSO, J. Climate, 19, 698–722, doi:10.1175/JCLI3631.1, 2006. 6089

Xie, S.-P., Deser, C., Vecchi, G. A., Ma, J., Teng, H., and Wittenberg, A. T.: Global warming pattern formation: Sea surface temperature and rainfall, J. Climate, 23, 966–986, doi:10.1175/2009JCLI3329.1, 2010. 6089

Yamaguchi, D. K.: A simple method for cross-dating increment cores from living trees, Can. J. Forest Res., 21, 414–416, 1991. 6080

Zinke, J., Dullo, W.-C., Heiss, G. A., and Eisenhauer, A.: ENSO and Indian Ocean subtropical dipole variability is recorded in a coral record off southwest Madagascar for the period 1659 to 1995, Earth Planet. Sc. Lett., 228, 177–194, doi:10.1016/j.epsl.2004.09.028, 2004. 6111

6110

Table 1. List of coral locations used for the EOF analysis of Fig. 3. The records were obtained from the database of Emile-Geay and Eshleman (2013) after selecting datasets that record $\delta^{18}O$ and cover the period 1860 to 1980.

	Site	Measurement	Latitude	Longitude	Range	Reference
1	Abrolhos	$\delta^{18}O$	28.46° S	113.77° E	1794–1994	Kuhnert et al. (1999)
2	Amedee Light.	$\delta^{18}O$	22.48° S	166.45° E	1660–1993	Quinn et al. (1998)
3	Bunaken	$\delta^{18}O$	1.50° N	124.83° E	1860–1990	Charles et al. (2003)
4	Clipperton Atoll	$\delta^{18}O$	10.30° N	109.00° W	1893–1994	Linsley et al. (2000)
5	Guam	$\delta^{18}O$	13.00° N	145.00° E	1790–2000	Asami et al. (2005)
6	Ifaty	$\delta^{18}O$	23.14° S	43.58° E	1659–1995	Zinke et al. (2004)
7	La Reunion	$\delta^{18}O$	21.03° S	55.25° E	1832–1995	Pfeiffer et al. (2004)
8	Laing	$\delta^{18}O$	4.15° S	144.88° E	1884–1993	Tudhope et al. (2001)
9	Lombok Strait	$\delta^{18}O$	8.25° S	115.50° E	1782–1990	Charles et al. (2003)
10	Madang	$\delta^{18}O$	5.22° S	145.82° E	1880–1993	Tudhope et al. (2001)
11	Mafia	$\delta^{18}O$	8.02° S	39.50° E	1622–1722	Damassa et al. (2006)
12	Mahe	$\delta^{18}O$	4.62° S	55.00° E	1846–1995	Charles et al. (1997)
13	Maiana Atoll	$\delta^{18}O$	1.00° N	173.00° E	1840–1994	Urban et al. (2000)
14	Mentawai	$\delta^{18}O$	0.13° S	98.52° E	1858–1997	Abram et al. (2008)
15	Moorea	$\delta^{18}O$	17.50° S	150.00° W	1852–1990	Boiseau et al. (1998)
16	Ningaloo Reef	$\delta^{18}O$	21.91° S	113.97° E	1878–1995	Kuhnert et al. (2000)
17	Palmyra	$\delta^{18}O$	5.88° N	162.00° W	1886–1998	Cobb et al. (2001)
18	Rabaul2006	$\delta^{18}O$	4.19° S	151.98° E	1867–1997	Quinn et al. (2006)
19	Rarotonga(2R)	$\delta^{18}O$	21.24° S	159.83° E	1726–1996	Linsley et al. (2008)
20	Rarotonga(3R)	$\delta^{18}O$	21.24° S	159.83° E	1874–2000	Linsley et al. (2006)
21	Ras Umm Sidd	$\delta^{18}O$	27.85° N	34.31° E	1751–1995	Felis et al. (2000)
22	Secas	$\delta^{18}O$	7.98° N	82.00° W	1707–1984	Linsley et al. (1994)

Table 2. Complexity as a function of θ and the matrix dimension np. Infinity symbols denote numbers that exceed machine capacity.

$np|\theta$	0.01	0.03	0.05	0.07	0.09
50	2×10^2	4×10^5	4×10^8	10^{11}	3×10^{13}
100	5×10^4	1×10^{12}	1×10^{18}	3×10^{23}	2×10^{28}
1000	9×10^{53}	1×10^{132}	5×10^{195}	4×10^{250}	3×10^{299}
5000	1×10^{277}	∞	∞	∞	∞

Table 3. Time-uncertain ensemble 95 % CI width means μ and standard deviations σ, and probabilities of the ensemble CI covering the original data with respect to θ.

θ	0.01	0.03	0.05	0.07	0.09
CI width, μ	0.38	0.47	0.50	0.53	0.55
CI width, σ	0.18	0.17	0.17	0.16	0.17
Coverage rate	0.97	0.99	0.97	0.98	0.98

6113

Fig. 1. Ensemble realization (top) and Power spectra (bottom) of age-perturbed coral data for **(a, b)** Isdale et al. (1998) (Havannah Island) and on the next page **(c, d)** Palmyra island δ^{18}O record from 1636 to 1703 (Cobb et al., 2003). The original data and corresponding spectral estimates are shown in red and the 95 % confidence intervals (CIs) and the 25–75 % interquartile range (IQR) of the perturbed data ensemble and corresponding spectra are shown by the shaded areas. Power-spectra are represented in variance-preserving coordinates (that is, the area under the curve is an exact measure of the variance contained in each frequency band).

6114

c) Palmyra island age-perturbed ensemble

d) Palmyra island (1636-1703) spectral estimate

Fig. 1. Continued.

6115

Palmyra age-correction: SB3, SB8 and SB13

SB3 ρ_0 = 0.22 ρ = -0.65
SB8 ρ_0 = -0.21 ρ = -0.59
SB13 ρ_0 = -0.14 ρ = -0.63

North American Drought Atlas (NADA) PC1

Fig. 2. Palmyra relict coral $\delta^{18}O$ records (top): published data SB3 in red, SB8 in orange and SB13 in brown (Cobb et al., 2003), the shaded area is the 95 % confidence interval of the ensemble of realizations with final time shifted $t_f \pm 10$ yr and $\theta_1 = \theta_2 = 0.05$ and the blue chronology corresponds to the realizations that maximize the correlation with the first principal component of the NADA record (bottom) over each continuous section. The correlation coefficients with NADA PC1 are also shown, wherein ρ is computed using the aligned Palmyra chronologies, and ρ_0 using the original chronologies.

6116

a) ENSO mode pseudocoral EOF - 25% variance

	EOF > 0		EOF < 0
	median		median
	95% quantile		95% quantile

Contours: SST ($^{\circ}$C) regressed onto ENSO mode

b) ENSO mode PC timeseries

c) ENSO mode PC spectrum
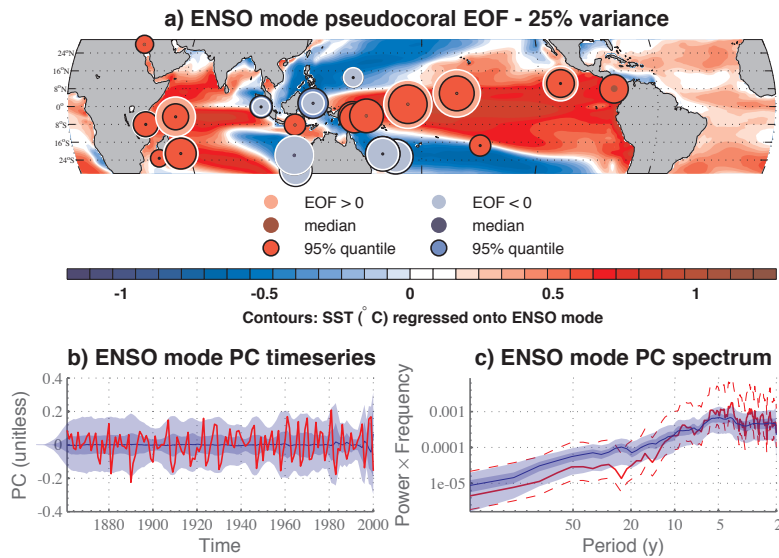
Fig. 3. Spatiotemporal uncertainty quantification on a pseudocoral network. **(a)** EOF loadings (circles) corresponding to the ENSO mode of an ensemble of age-perturbed pseudocoral records with miscounting rate $\theta = 0.05$. EOF loadings for error-free data are shown in light colors circled in white, while the median and 95 % quantile are shown by dark disks and black-circled disks, respectively. Contours depict the SST field associated with the mode's principal component PC **(b)**, whose power spectrum is shown in **(c)**. Results for the time-uncertain ensemble are shown in blue: median (solid line), 95 % confidence interval (light-filled area) and interquartile range [25–75 %] (dark-filled area). Results for the original (error-free) dataset are depicted by solid red lines. Dashed red lines denote χ^2 error estimates for the MTM spectrum of the error-free dataset.

6117

a) Original signals

b) Age-perturbed signals

c) Layer-counted signals
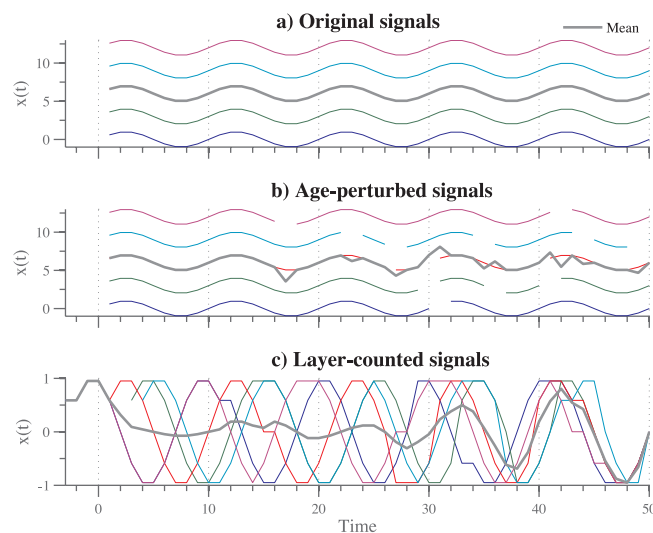
Fig. 4. Illustration of the phase distortion of coherent signals due to miscounted annual bands ($\theta_1 = \theta_2 = 0.05$). (top) Five hypothetical harmonic signals (middle) the same signals after introduction of age perturbations (bottom) the same age-perturbed signals, after making the (erroneous) assumption that 1 layer amounts to 1 yr. Note the loss of coherency reflected in the ensemble average, which is performed as part of composites or most climate reconstructions.
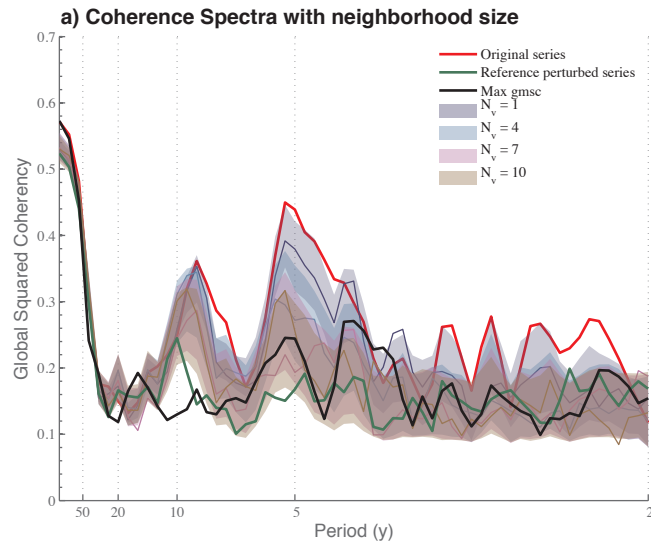
6118

Fig. 5. Spectral density of the ENSO mode PC for $\theta = 0.001$. Compare to Fig. 3, bottom right.

Fig. 6. Global coherence spectra for time-uncertain ensembles generated with $\theta_1 = \theta_2 = 0.001$ (orange), $\theta_1 = \theta_2 = 0.01$ (dark blue) and $\theta_1 = \theta_2 = 0.05$ (gray). The ensemble medians are illustrated in dashed lines and their corresponding colored patches represent the ensemble 95 % CIs. The thick red curve is the coherence spectrum of the error-free time series.

a) Coherence Spectra with neighborhood size

Fig. 7. Coherence spectra **(a)** and on the next page Hamming distance distributions to the original chronology **(b)** for ensembles of chronologies sampled around the reference perturbation matrix (with $\theta = 0.05$), in restricted neighborhood of sizes $2 \cdot N_v + 1$ for N_v increasing from 1 to 10. The thick red curve represents the coherence spectrum for the error-free data, the results related to the reference perturbed dataset are shown in green and plotted in black are the results for the ensemble realization (sampled across the entire solution space) that maximizes the GMSC (as described in Sect. 4.1). Finally the results for the neighborhood ensembles (shaded areas): the 95 % CIs and Hamming distance distributions, and corresponding realizations maximizing the GMSC (solid lines) are shown in dark blue, blue, pink and light brown for $N_v = 1$, 4, 7 and 10, respectively.
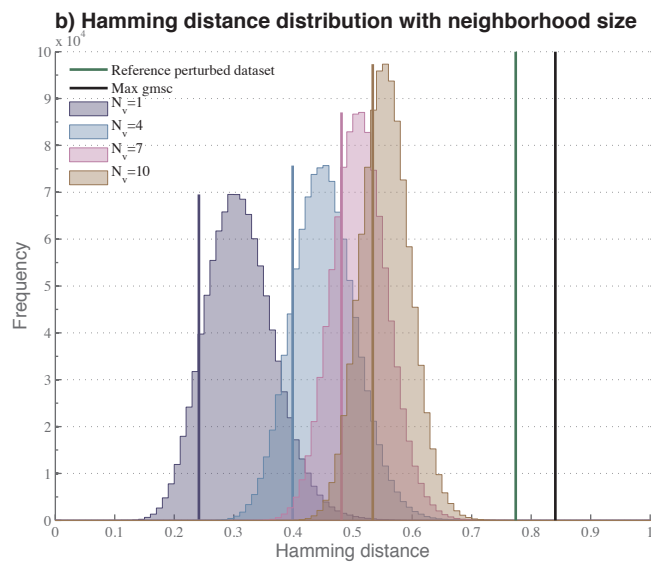
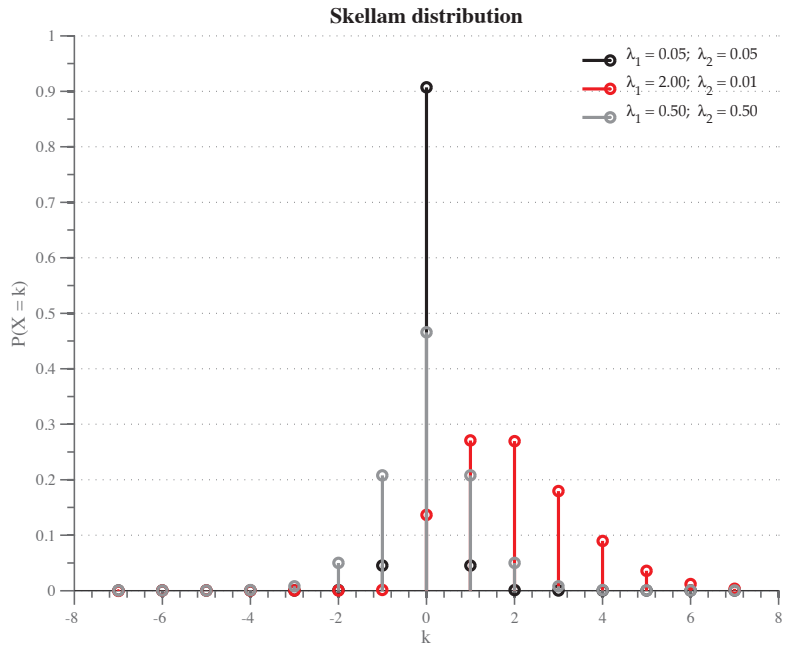### b) Hamming distance distribution with neighborhood size

Fig. 7. Continued.

Fig. A1. Skellam probability mass functions for different values of its parameters λ_1 and λ_2.