Correlation-based interpretations of paleoclimate data – where statistics meet past climates

Jun Hu^a, Julien Emile-Geay^{a,*}, Judson Partin^b

^a*Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089, USA*

^b*Institute for Geophysics, University of Texas at Austin, Austin, TX 78758, USA*

Abstract

Correlation analysis is omnipresent in paleoclimatology, and often serves to support the proposed climatic interpretation of a given proxy record. However, this analysis presents several statistical challenges, each of which is sufficient to nullify the interpretation: the loss of degrees of freedom due to serial correlation, the test multiplicity problem in connection with a climate field, and the presence of age uncertainties. While these issues have long been known to statisticians, they are not widely appreciated by the wider paleoclimate community; yet they can have a first-order impact on scientific conclusions. Here we use three examples from the recent paleoclimate literature to highlight how spurious correlations affect the published interpretations of paleoclimate proxies, and suggest that future studies should address these issues to strengthen their conclusions. In some cases, correlations that were previously claimed to be significant are found insignificant, thereby challenging published interpretations. In other cases, minor adjustments can be made to safeguard against these concerns. Because such problems arise so commonly with paleoclimate data, we provide open-source code to address them. Ultimately, we conclude that statistics alone cannot ground-truth a proxy, and recommend establishing a mechanistic understanding of a proxy signal as a sounder basis for interpretation.

*Corresponding author

Email address: julieneg@usc.edu (Julien Emile-Geay)

Keywords: speleothems; calibration; age-uncertainties; false discovery rate; spurious significance

1. Introduction

Inferring past climate conditions from proxy archives is a central tenet of paleoclimatology. The calibration of paleoclimate proxies is accomplished in two main ways: space-based calibrations and time-based calibrations (defined below). In space-based calibrations, the values of a proxy at different locations are calibrated to measured climate indicators at the same locations, as exemplified by the calibration of paleothermometers in the core-top of marine sediments (e.g. Tierney and Tingley, 2014; Khider et al., 2015). This approach is relatively forgiving of time uncertainties, as long as core-top values are broadly contemporaneous, in relation to the question being asked of the cores. In time-based calibrations, on the other hand, proxy timeseries overlapping with the instrumental era are calibrated against an instrumental target (e.g. Jones et al., 2009; Tingley et al., 2012), via correlation analysis or the closely-related linear regression.

Thus “ground-truthing” a proxy record often involves establishing that its correlation to an instrumental climate variable (whether local, regional, or global) is significant in some way. Significance of correlations is most commonly assessed via a t -test, which assumes that samples are independent, identically-distributed, and Gaussian. However, these criteria may not be fulfilled in paleoclimate timeseries due to their intrinsic properties (Ghil et al., 2002).

Indeed, the loss of degrees of freedom due to autocorrelation has long been known to challenge the assumption of independence (Yule, 1926), though workarounds are known (e.g. Dawdy and Matalas, 1964). Non-Gaussianity may also prove an issue, especially for precipitation timeseries, though relatively simple transformations may alleviate it (Emile-Geay and Tingley, 2016).

Additionally, correlating proxies with instrumental climate fields is a common way of establishing the ability of a proxy to capture large-scale climate

information. Unfortunately when implemented as a mining exercise using a large, spatially gridded dataset, test multiplicity becomes a problem. We will review how this problem may be successfully circumvented using simple statistical approaches (Benjamini and Hochberg, 1995; Storey, 2002).

Finally, the presence of age uncertainties may bring substantial uncertainties to time-based correlations between records (e.g. Crowley, 1999; Wunsch, 2003; Black et al., 2016). We will show a robust approach to quantifying age uncertainties and how they propagate to correlation and other analyses.

The article is structured as follows. In Section 2 we show the importance of considering autocorrelation in cross-correlation analyses. In Section 3, we briefly introduce the "multiple hypothesis test" problem and the false discovery rate and show how it affects correlations with a climate field. In Section 4, we introduce the effects of age uncertainties, how they influence the interpretation of a speleothem record, and how this compounds with the other two challenges. We finish with a discussion of the significance of these results, and propose strategies to mitigate these statistical issues going forward.

2. Challenge #1: serial correlation

2.1. Theory

The most common way to determine the significance of Pearson's product-moment correlation involves a t -test. Student's t distribution is fully determined by the number of degrees of freedom available in the sample (ν). For N independent samples, $\nu = N - 2$, but it may be considerably lower when this assumption is violated, leading to overconfident assessments of significance.

As an example, consider correlations between two timeseries $x(t)$ and $y(t)$ generated by autoregressive processes of order 1 (a common timeseries model for serially correlated data (e.g. Emile-Geay, 2016, Chapter 8)). Each process is evenly sampled 500 times and their correlation coefficient is 0.13, which is significant at the 5% level assuming independence (hence, $\nu = 498$). However,

the lag-1 autocorrelation of each time series (ϕ) is 0.8, which is common for climate variables like temperature, as well as for many paleoclimate records, which tend to have a red spectrum (Ghil et al., 2002). This means that neighboring samples are highly dependent, so the effective numbers of degrees of freedom, ν_{eff} , is much lower. This number may be estimated via the following relation (Dawdy and Matalas, 1964):

$$\nu_{\text{eff}} = N \frac{1 - \phi_x \cdot \phi_y}{1 + \phi_x \cdot \phi_y} \quad (1)$$

where ϕ_x , ϕ_y are the lag-1 autocorrelation coefficients of two time series x , y respectively.

Based on equation 1, when either lag-1 autocorrelation coefficient increases, the effective number of degrees of freedom decreases, and the p -value of the test increases. In this case, the effective number of degrees of freedom decreases from 498 to 99 after considering the autocorrelation, and the p -value rises to 0.19, suggesting the correlation is no longer significant at the 5% level. Fig. 1 shows how the p -value and the degrees of freedom change for a time series of 500 samples and a fixed correlation of 0.13 just by changing the autocorrelation coefficients ($\phi_x = \phi_y = \phi$ for simplicity). As the autocorrelation increases, the p -values increases, and the degrees of freedom decrease. When all samples are independent ($\phi_x = \phi_y = 0$), the p -value is far smaller than 5%. When the autocorrelation increases to about 0.65, the p -value becomes larger than 5%, making the correlation insignificant at this level. The problem only worsens as ϕ increases, and as we shall see in this article, values above 0.8 are quite typical of paleoclimate timeseries.

Autocorrelation is thus a very serious challenge, which alone can substantially raise the bar of a significance test; if ignored, it may lead to overconfident assessments of significance.

2.2. Application

To see this effect at work in the real world, consider the example of Proctor et al. (2000), who used the band width in a stalagmite (SU-96-7) from Uamh

Figure 1: The p -value and numbers of degrees of freedom (DOF) of the correlation (0.13) between two AR(1) time series (500 samples each) with the changing autocorrelation ϕ . The green dashed line is the 5% criteria for 5% level significance test.

84 an Tartair (northwest Scotland) to reconstruct the North Atlantic Oscillation
 85 (NAO). The record was dated by counting annual bands, with only 17 bands
 86 as double annual bands, implying a counting error less than 20 years. When
 87 compared to the whole length of the entire 1087 year-long record, this amounts
 88 to only 2%. Therefore, the influence of age uncertainties can be neglected to
 89 first order.

90 The climatic interpretation of the stalagmite was based on the high correla-
 91 tion between the band width and the temperature/precipitation ratio ($r = 0.80$)
 92 as well as the correlation between band width and the winter NAO index
 93 ($r = -0.70$) by using decadal-smoothed data. Here we apply the effective
 94 degrees of freedom in testing the significance of correlation, since the correlation
 95 significance may be biased by autocorrelation due to the effect of smoothing.
 96 Also, inherent aspects of these records leads to complications using statistics
 97 based on normally distributed populations, as the band width distribution of
 98 the stalagmite record is bimodal instead of normal. The t -test for correlation
 99 significance assumes that both time series are normally distributed, negating its

100 use as a statistical tool unless appropriate transformations are made.

101 Considering the autocorrelation of the smoothed data, the high correlation
102 between the band width of stalagmites and the ratio temperature/precipitation
103 (T/P) in the instrumental period is not significant at 5% significance level (the
104 adjusted p -value is 0.44). The correlation between the band width of stalagmites
105 and winter NAO is also not significant, because of high autocorrelations of the
106 smoothed time series of the band width ($\phi = 0.99$), T/P ($\phi = 0.99$) and winter
107 NAO ($\phi = 0.95$). However, this result is based on an assumption of normality,
108 and as discussed above, the distribution of the band width in this speleothem
109 is bimodal, hence non-normal (not shown). Thus, transforming the non-normal
110 series to normality (Emile-Geay and Tingley, 2016) is necessary. After this
111 transformation, the correlations pass the significance test at the 5% level: for
112 the correlation between the band width and T/P, ν_{eff} is 93 ($N = 115$), and the
113 p -value is 3×10^{-3} ; for the correlation between the band width and winter NAO,
114 ν_{eff} is 95 ($N = 126$), and the p -value is 4×10^{-2} , just under the 5% threshold.
115 While we conclude that the original interpretation is supported by our analysis,
116 the authors reached this conclusion thanks to error compensation, potentially
117 undermining their point.

118 We note, however, that the decrease of DOF due to smoothing was consid-
119 ered when this reconstruction was used for studying the long-term variability of
120 the NAO in the high-profile study of Trouet et al. (2009).

121 **3. Challenge #2: Test multiplicity**

122 *3.1. Theory*

123 When assessing correlations with a field, multiple tests are carried out at
124 different locations simultaneously. For example, if a correlation test is applied
125 to 1000 locations with a significance level of 5%, one would expect about 50
126 hypotheses to be falsely rejected (in this case, 50 correlations would be deemed
127 significant when in fact they are not), which is unacceptably high. The funda-
128 mental problem is that the test level α (the probability of false positives, i.e. the

129 probability of falsely rejecting the null hypothesis of zero correlation) applies to
130 pairwise comparisons, but not to multiple such comparisons. The finer the grid,
131 the more such tests are simultaneously carried out, and the higher the risk of
132 identifying spurious correlations as “significant” when in fact they are not. This
133 test multiplicity problem is well known in the statistical literature and solutions
134 exist (Benjamini and Hochberg, 1995; Storey, 2002).

135 In particular, the False Discovery Rate (FDR) procedure (Benjamini and
136 Hochberg, 1995) has been widely applied ($> 34,000$ Google Scholar citations at
137 the time of writing) to control the proportion of falsely rejected null hypotheses
138 out of all rejected null hypotheses, thus offering a level of scientific rigor that
139 naive correlation testing does not afford. The term “false discovery” here is syn-
140 onymous with “falsely identified significant correlations”. Instead of restricting
141 the occurrence of falsely rejected null hypotheses, the FDR procedure allows and
142 controls the proportion of erroneously rejected null hypotheses. Setting $q = 5\%$
143 in the FDR procedure, guarantees that 5% or fewer of the locations where the
144 null hypothesis is rejected are false detections on average, so the proportion of
145 false rejections is controlled. The FDR procedure of Benjamini and Hochberg
146 (1995) proceeds as follows:

- 147 1. Carry out the test (calculate p -values) at all m locations;
- 148 2. Rank the p -values $p_{(i)}$ in increasing order: $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$;
- 149 3. Define k as the largest i for which $p_{(i)} \leq q \frac{i}{m}$;
- 150 4. Reject hypotheses at locations $i = 1, 2, \dots, k$.

151 In this procedure, the FDR treats locations with low p -values as most sig-
152 nificant, ranking p -values from high to low. If the largest p -value is less than its
153 corresponding threshold, then all tests are regarded as significant. If the largest
154 p -value is greater than its corresponding threshold, then it is compared to the
155 second largest p -value with a more restricted threshold – and so on. These
156 thresholds guarantee that the expected rate of falsely positive hypotheses are
157 smaller than α . Through this procedure, the number of rejected hypotheses is k ,
158 and the expected number of falsely rejected hypotheses is smaller than $mp_{(k)}$,

159 such that the fraction of falsely rejected hypotheses is smaller than $mp_{(k)}/k$.
 160 The third step in the FDR procedure limits the fraction of falsely rejected hy-
 161 potheses to be smaller than q , which is the threshold for the fraction of falsely
 162 rejected hypotheses. Thus the FDR procedure ensures the fraction of erro-
 163 neously detected significant locations is smaller than a specified threshold.

Figure 2: Illustration of the traditional significance test procedure and the FDR procedure on an illustrative example, with $q = \alpha = 20\%$. p -values at each grid point ($p_{(i)}$) are ranked in increasing order, plotted against i/m , where m is the total number of grid points. The blue dashed line is the traditional α threshold for the p -value. Green dots indicate they are significant by both traditional and FDR procedure, and blue dots indicate they are only significant by the traditional procedure.

164 This procedure is graphically illustrated by Fig. 2, adapted from Ventura
 165 et al. (2004). In order to clearly show the difference between the traditional and
 166 the FDR significance test procedure, q and α were both set to 20%. At each
 167 location ($p_{(i)}$), p -values were ranked in increasing order and plotted against i/m ,
 168 where m is the total number of locations. The blue dashed line is the traditional
 169 significance threshold. All dots below the blue dashed line are significant by the
 170 traditional procedure. With the FDR procedure, only those p -values under the
 171 red line (green dots) are significant. Therefore, some p -values are deemed sig-
 172 nificant by the traditional procedure but not by the FDR procedure. Applying

173 this methodology lowers the likelihood of identifying spurious correlations as
 174 significant, hence making correlation tests more stringent. We should note that
 175 the absence of significant correlations in one test does not imply that the cor-
 176 relation is inexistent – only that the data do not provide enough evidence to
 177 reject the null hypothesis of zero correlation.

178 *3.2. Application*

179 Ventura et al. (2004) proposed a simple implementation of the FDR pro-
 180 cedure for climate fields, and showed that its assumptions are robust to the
 181 spatial correlation levels typical of climate fields. Here we use the code of Ben-
 182 jamini and Hochberg (1995), which is equivalent. To show how to apply this
 183 procedure to field correlations, we use the study of Zhu et al. (2012) as an exam-
 184 ple. The authors generated a cellulose $\delta^{18}\text{O}$ record of Merkus pines for the past
 185 140 years in Kririrom National Park, southern Cambodia (KRPM15B, 11.29°N;
 186 104.25°E; 675 m). This record was dated by ring-counting. The authors assert
 187 that this cellulose $\delta^{18}\text{O}$ record is dominantly controlled by convection over the
 188 Indo-Pacific warm pool (IPWP), an interpretation buttressed by high correla-
 189 tions of cellulose $\delta^{18}\text{O}$ with instrumental precipitation and outgoing longwave
 190 radiation over the IPWP (Fig. 7 in Zhu et al. (2012)). This explanation is
 191 reasonable because the cellulose $\delta^{18}\text{O}$ is mainly controlled by the $\delta^{18}\text{O}$ in pre-
 192 cipitation during the rainy season, which is often depleted when the rainfall
 193 increases (the so-called “amount effect”; Dansgaard, 1964). During El Niño
 194 events, the precipitation in Southeast Asia is usually suppressed, which would
 195 lead to higher $\delta^{18}\text{O}$ values.

196 Here we consider the false discovery rate in this spatial correlation (Fig. 7
 197 in Zhu et al. (2012)). When the FDR is considered, none of the correlations are
 198 found significant, even at the relatively permissive 10% level chosen by authors
 199 (Fig. 3). This calls into question the proposed relationship between cellu-
 200 lose $\delta^{18}\text{O}$ and interannual changes in tropical convection at this site. However,
 201 we did find that the correlation between the cellulose $\delta^{18}\text{O}$ and (unsmoothed)
 202 NINO4 index significant at the 5% level ($r = 0.45$, $p\text{-value} = 2.3 \times 10^{-6}$), con-

203 sidering autocorrelation as above. This indicates that cellulose $\delta^{18}\text{O}$, or at
 204 least rainfall $\delta^{18}\text{O}$ at the site, may be connected with large scale changes due
 205 to the El Niño-Southern Oscillation. However, the explanation that local pre-
 206 cipitation changes during El Niño events dominate the cellulose $\delta^{18}\text{O}$ may need
 207 further examination, for instance via forward modeling (Evans, 2007).

Figure 3: Spatial correlation of October Kirirom cellulose $\delta^{18}\text{O}$ values with the October-
 November-December mean of (a) CMAP precipitation, and (b) NOAA interpolated OLR.
 Black dots indicate that the correlation does not pass the significance test at the 90% level

208 4. Challenge #3: Age uncertainties

209 4.1. Theory

210 Age uncertainties have long been known to affect inferences made from pa-
211 leoclimate data (e.g. Clark and Thompson, 1979), including correlation analysis
212 (see Mudelsee, 2001; Rehfeld and Kurths, 2014, for recent examples). Indeed,
213 Wunsch (2003) showed that even two randomly generated, unrelated time se-
214 ries could appear to correlate well with each other by adjusting the chronology
215 within age uncertainties.

216 Thus, quantifying age uncertainties is critical to the analysis of paleoclimate
217 records. Many methods have been developed to do so (e.g. Haslett and Parnell,
218 2008; Blaauw and Christen, 2011). Additionally, various methods have been
219 devised to deal with correlation (or covariance) under age uncertainties (Haam
220 and Huybers, 2010).

221 In the next section, we illustrate how age uncertainties may influence cor-
222 relation analysis, and how they may compound the effects of autocorrelation
223 and test multiplicity by drawing an example from a high-profile publication
224 (McCabe-Glynn et al., 2013).

225 4.2. Compound challenges in one: the case of Crystal Cave

226 Recently, McCabe-Glynn et al. (2013) applied time-based calibrations to
227 $\delta^{18}\text{O}$ data from a stalagmite from Crystal Cave, southern California. They
228 interpreted the record as a proxy for sea surface temperature (SST) in the
229 Kuroshio Extension region, arguing that warm SST in the area may generate
230 southwesterly wind anomalies over southern California, bringing isotopically-
231 enriched moisture to the cave. They also found a strong 22-year periodicity,
232 which they linked to solar cycles, and found that some southwestern North
233 American droughts coincided with episodes of warm Kuroshio Extension SSTs,
234 as reconstructed from the Crystal Cave stalagmite. Some of these conclusions
235 are based on a correlation analysis that encounters the three challenges of in-
236 terest in this study: serial correlation, test multiplicity and age uncertainties.
237 Here we will show one way to properly address these challenges.

4.3. Effect of serial correlation

We reuse the $\delta^{18}\text{O}$ data from stalagmite CRC-3, collected from Crystal Cave in Sequoia National Park, California (36.59°N; 118.82°W; 1,386 m), which was interpolated at annual scale and archived online¹. As in the original study, the record is correlated to SST anomaly data from the Kaplan SST v2 dataset (Kaplan et al., 1998) (1856-2007).

The correlation is shown in Fig. 4b, to be compared with Supplementary Fig. S6a in McCabe-Glynn et al. (2013). Because both the $\delta^{18}\text{O}$ series and SST field are intrinsically autocorrelated (the autocorrelation of $\delta^{18}\text{O}$ series is 0.95, and the autocorrelation of SST is 0.11-0.76, depending on location), the effective number of degrees of freedom is much lower than its theoretical value ($N - 2 = 149$). Indeed ν_{eff} is less than 60 in the Kuroshio Extension region (Fig. 4a). Hence, when this effect is considered (Fig. 4c), far fewer correlations pass the significance test (Fig. 4b,c). While McCabe-Glynn et al. (2013) had considered the effect of serial correlation using the method of Macias-Fauria et al. (2012), they did not graphically represent these results, giving little indication of where the relationship might be reliable. Nonetheless, our result is consistent with theirs, in that correlations over the Kuroshio Extension region pass the significance test with both approaches.

¹<ftp://ftp.ncdc.noaa.gov/pub/data/paleo/speleothem/northamerica/usa/california/crystal2013.txt>

Figure 4: (a) The effective number of degrees of freedom of the correlation between Crystal Cave $\delta^{18}\text{O}$ and SST from 1856-2007; The correlation between these series, considering the test multiplicity problem (False Discovery Rate) (d, e), or not (b, c); considering autocorrelation (c,e) or not (b,d). Black dots indicate that the correlation does not pass the significance test at the 5% level.

4.4. Effect of test multiplicity

Since the correlation between the $\delta^{18}\text{O}$ record and SST is also a field correlation, we need to consider test multiplicity. The result is shown in Fig. 4d,e. If autocorrelation is ignored, the FDR procedure results in fewer correlations passing the significance test (Fig. 4d vs Fig. 4b). When both autocorrelation and multiplicity are considered, no correlation passes the significance test (Fig. 4e).

263 This result suggests that the correlation between $\delta^{18}\text{O}$ record and instrumental
264 SST may not be used as a basis for the record's interpretation.

265 4.5. *Effect of age uncertainties*

266 Another problem compounding serial correlation stems from the fact that the
267 age model for the speleothem carries uncertainties of years to decades (Cheng
268 et al., 2013), and these uncertainties may propagate to other inferences made
269 from the proxy. The age uncertainties were quantified in McCabe-Glynn et al.
270 (2013) using the StalAge algorithm (Scholz and Hoffmann, 2011). While the Sta-
271 lAge code exports 95%-confidence limits for the corresponding ages, it does not
272 export possible age ensembles, which are essential for propagating uncertainty
273 to other inferences. Here we leverage the power of ensembles to quantifying age
274 uncertainties in correlation analysis.

275 4.5.1. *Age model*

276 We chose to model age uncertainties using Bchron (Haslett and Parnell,
277 2008), a Bayesian probability model allowing for random variations in accu-
278 mulation rate between tie points. Bchron is capable of dealing with outliers
279 and hiatuses (Parnell et al., 2011), and naturally produces more hiatuses than
280 StalAge. Bchron proceeds as follows: for each dated sample, the algorithm ran-
281 domly selects a calendar date consistent with the age information (measured
282 age and age error) and monotonicity (deeper sample with older age). It then
283 inserts several points (at depths we want to know ages) between dated sam-
284 ples consistent with monotonicity, and then linearly interpolates between those
285 points. Finally, it repeats this process many times until enough realizations fit
286 the measured ages. The main advantage of Bchron here is the ability to extract
287 an ensemble of age models all consistent with the posterior distribution of ages.
288 For a review of different approaches to age modeling, see Scholz et al. (2012).

289 The age models are compared in Fig. 5, and one can see that they are quite
290 close. The choice of age model (blue vs black curve) introduces relatively small
291 differences in the median age model, but the inclusion of a full ensemble of 1000

Figure 5: The age modeling results of the Crystal cave $\delta^{18}\text{O}$ record using a Bchron age model. The gray area is the 95% confidence interval of the age at each depth. The red lines show 10 random paths out of the 1,000 age models generated, and the blue curve shows the StalAge-generated model used by McCabe-Glynn et al. (2013)

292 plausible age realizations makes a great difference indeed. Each of these age
 293 realizations corresponds to a different $\delta^{18}\text{O}$ time series: Fig. 6 (left) shows three
 294 of them, corresponding to the lower (2.5%), median (50%) and upper (97.5%)
 295 quantiles of the age distribution. One can see that many of the major features
 296 of the $\delta^{18}\text{O}$ timeseries can shift by about 50 years, making a correlation to
 297 instrumental data fraught with uncertainty.

298 This may be seen in more detail on Fig. 5. While the median age models
 299 from the two methods (blue and black curves) do appear quite close, there are
 300 large offsets between $\delta^{18}\text{O}$ timeseries (Fig. 6, right), especially before AD 1960.
 301 For instance, the large peak ca 1890 in McCabe-Glynn et al. (2013), is centered
 302 around 1900 in the median Bchron age model. Such differences are especially
 303 significant if one tries to correlate them to other climatic timeseries, as we now
 304 do.

Figure 6: Left panel: The median, 97.5% quantile and 2.5% quantile of the ensemble of 1000 Crystal $\delta^{18}\text{O}$ time generated from the Bchron model. Right panel: The time series of $\delta^{18}\text{O}$ record from McCabe-Glynn et al. (2013) (blue) and the median of $\delta^{18}\text{O}$ record from the Bchron model (black).

4.5.2. Correlations considering age uncertainty and autocorrelation

We assess the impact of chronological uncertainties by repeating the previous analysis on each of the 1,000 realizations of the $\delta^{18}\text{O}$ time series, similarly interpolated to an annual scale to facilitate correlations to SST. In the following analysis, we will also take autocorrelation into account.

Using the full age ensemble, one obtains 1000 $\delta^{18}\text{O}$ -SST correlations for each grid point, from which one may infer an empirical distribution, whose median, 2.5% and 97.5% percentiles, and interquartile range (IQR) are reported in Fig. 7. The median is the aspect of the distribution that most studies use exclusively (for Gaussians, the median, mean and mode coincide). Due to the construction of our ensemble, the 2.5% quantile gathers some of the strongest negative correlations, and the 97.5% quantile gathers some of the strongest positive correlations. The IQR measures the spread between the 25% and 75% quantiles (the width of the distribution), and is therefore an indication of the spread of correlations due to age uncertainties alone.

The pattern of correlations for the median age model is similar to Fig. 4b, but the absolute values of correlations are much smaller, and the positive center in the North Pacific is shifted southward. Also, none of the median correlations

Figure 7: Correlations considering age uncertainty. The median (a), interquartile range (b), the 2.5% (c) and the 97.5% (d) percentile of the correlation between the $\delta^{18}\text{O}$ age ensemble and SST. Black dots indicate that the correlation does not pass the significance test at the 5% level, accounting for serial correlation. Note that the interquartile range here is a measure of distributional spread, and has no measure of significance attached to it.

pass the significance test at the 5% level. Since McCabe-Glynn et al. (2013) use the StalAge model, this suggests that the correlation between the $\delta^{18}\text{O}$ record and instrumental SST may be dependent on the age model. However, Fig. 5 clearly shows that the StalAge model is within the 95% confidence bounds of the Bchron age model, which underlines that age uncertainties are generally quite large compared to the timescale of variability in the SST record, and should therefore be accounted for in the analysis of correlations.

The pattern of the range of the correlation (IQR, Fig. 7b) is quite similar to Fig. 4b, which indicates that the regions of highest correlation in McCabe-Glynn

332 et al. (2013) correspond to the regions of largest uncertainties in the age models.
 333 The 97.5% quantile and the 2.5% quantiles (Fig. 7c,d) also correspond to the
 334 regions of positive/negative correlation in Fig. 4b, and the correlation in some
 335 regions passes the significance test. However, the corresponding pattern differs
 336 from age ensemble to age ensemble, and from the published result, indicating a
 337 lack of robust relationship on which to build a reliable interpretation.

338 *4.6. Correlations considering all three challenges*

339 Adding to these challenges, when the multiple-test problem is considered,
 340 none of the correlations for the 97.5% or 2.5% quantiles of the age ensemble pass
 341 the significance test (Fig. 8). This is because, while some p -values fall below
 342 the nominal threshold (Fig. 9, blue dashed lines), none drop under the FDR
 343 threshold (Fig. 9, red line), implying that these correlations are not significant
 344 at the 5% level. Thus the interpretation of Crystal Cave $\delta^{18}\text{O}$ as a proxy for
 345 SST in the Kuroshio Extension (or SST anywhere) should be revisited.

Figure 8: Same as Fig. 7, but testing for field correlations while controlling for the False Discovery Rate.

4.7. Time-uncertainty spectral analysis

Finally, we assess the influence of chronological uncertainties on inferences made in the spectral domain. We use the multi-taper method (Thomson, 1982), which achieves an optimal tradeoff between leakage and resolution. We use a half-bandwidth parameter of 4, which favors statistical significance over resolution, and is therefore the most conservative setting. The multi-taper spectra of the $\delta^{18}\text{O}$ record (855-2007AD) from the two age models are presented in Fig. 10. It shows that the $\delta^{18}\text{O}$ record derived from the median of the Bchron ensemble has a dominant period ca 18 years while the published data exhibit a dominant period of 21 years. This is clearly within age uncertainties, despite McCabe-Glynn et al. (2013) having used REDFIT, a variant of the LombScargle

Figure 9: Illustration of the FDR procedure on the 2.5% quantile (a) and the 97.5% quantile (b) correlations shown in Fig. 8, using the false discovery rate $q = 5\%$ (red line). p -values at each grid point ($p_{(i)}$) are ranked in increasing order, plotted against i/m , where m is the total number of grid points. The blue dashed line is the traditional 5% threshold for the p -value. Dots with p -values below this threshold are shown in blue. In this example, many dots fall below the nominal threshold (5%), but none fall below the red line, which means that they are not significant according to the FDR-controlling procedure.

periodogram (Schulz and Mudelsee, 2002). Therefore, the multi-taper spectral method is suitable for the comparison. Considering that much of the uncertainty lies above the red line and gray area in Fig. 10, it is hard to distinguish significant periodicities from red noise. This suggests that there may not be any notable harmonic cycles in the $\delta^{18}\text{O}$ record at any scale less than 1000 years. Thus more evidence would be needed to draw a connection to the 22-year Hale solar cycle, as done in McCabe-Glynn et al. (2013).

We showed how to address the three main challenges of correlation analysis and how these challenges weaken the conclusions of the original study. The example illustrates how each challenge in isolation would be enough to question the main conclusions, and their combination even more so. In closing, we note that although we do not find significant correlations between the Crystal Cave record and sea-surface temperature, we cannot rule out the existence of such correlations. The short calibration period and relatively large age uncertainties simply preclude the establishment of significant correlations with the current

Figure 10: The MTM-estimated spectra of the $\delta^{18}\text{O}$ record from the Bchron age model (blue line and gray shaded area) compared to the spectrum of the published record, together with a simulated AR(1) benchmark.

372 datasets.

373 Finally, we stress that the interpretation of the oxygen isotope composition
 374 of speleothem calcite is complex (LeGrande and Schmidt, 2009), and that other
 375 factors may influence the $\delta^{18}\text{O}$ record in Crystal Cave. The variation of $\delta^{18}\text{O}$
 376 of precipitation in California can be affected by changes in condensation height
 377 (Buenning et al., 2013), and $\delta^{18}\text{O}$ in precipitation is known to be influenced
 378 by source-water composition due to shifts in storm-track location (Oster et al.,
 379 2015). Thus, the published interpretation of Crystal Cave $\delta^{18}\text{O}$ needs to be
 380 further investigated.

381 5. Discussion

382 In a provocative paper, Ioannidis (2005) concluded that “most published re-
 383 search findings are false”. Central to this problem is a widespread tendency to
 384 hunt for statistical effects (often, correlations) that pass a significance thresh-
 385 old of 95% (i.e. a 5% probability of false positive), a practice often dubbed

386 “*p*-hacking”. While the paper focused on biomedical research, some of its con-
387 clusions transfer to paleoclimatology, and the Earth sciences at large. In par-
388 ticular, the existence of important challenges in correlation analysis (autocor-
389 relation, false-discovery rate and age uncertainties) should be recognized by all
390 practicing paleoclimatologists. In this article we illustrate these challenges using
391 three examples, showing how published interpretations may be changed by one
392 challenge alone, or by a combination of them.

393 We began with showing how autocorrelation and non-normality affect cor-
394 relation analysis by using the example from Proctor et al. (2000), who recon-
395 structed the NAO index using layer thickness in a stalagmite from Scotland.
396 We found that properly taking those into account did not change the interpre-
397 tation, though the latter would have been more robust had the authors taken
398 these challenges into consideration. Autocorrelation is commonplace in paleo-
399 climate proxies because many processes (e.g. bioturbation in sediment cores,
400 groundwater mixing in speleothems, and firn diffusion in ice cores) smooth out
401 climate signals. Autocorrelation should always be a concern unless it can be
402 demonstrated otherwise.

403 The test multiplicity problem is another serious challenge of correlation anal-
404 ysis, as we showed in the example of Zhu et al. (2012). The spatial correlations
405 between their cellulose $\delta^{18}\text{O}$ record and instrumental precipitation/outgoing
406 longwave radiation over the Indo-Pacific Warm Pool in this paper are not sig-
407 nificant after considering the false discovery rate. However, we find correlations
408 with the NINO4 index significant, with the index explaining about 25% of the
409 record’s variance.

410 Age uncertainties also challenge the robustness of correlation analysis and
411 we show that they may also combine with other challenges in the example of
412 McCabe-Glynn et al. (2013), who claimed, on the basis of correlations to the
413 SST field, to identify a relationship between the $\delta^{18}\text{O}$ record of a speleothem
414 from southern California and the SST in the Kuroshio Extension region. We
415 show, by considering all three challenges, that no correlation survives the test.

416 These three examples lead us to draw attention to the importance of:

- 417 • using established statistical procedures to guard against the misleading
418 impacts of spurious correlations. Several books (Wilks, 2011; Mudelsee,
419 2013; Emile-Geay, 2016) address these challenges in more detail.
- 420 • using the rich output of age modeling software (not just the median age)
421 to appraise the effect of age uncertainties on a study’s conclusions.
- 422 • establishing a mechanistic understanding for proxy signals, and only rely-
423 ing on statistical approaches when there are sufficient numbers of degrees
424 of freedom to unequivocally reject chance correlations.

425 While we do not dispute that the records scrutinized here contain potentially
426 valuable climatic information, our main message is that it is often impossible
427 to establish so by a purely statistical approach. Instead, we encourage detailed
428 process studies to elucidate the climatic and/or hydrological controls on $\delta^{18}\text{O}$,
429 or other proxies, in various archives.

430 For speleothems, possible strategies involve the forward modeling of cave
431 processes (such as Baker et al., 2012; Partin et al., 2013, and others), and/or
432 cave instrumentation (such as Spötl et al., 2005; Partin et al., 2012, and others)
433 to better ascertain the processes that control the recorded oxygen isotope signal.
434 For many speleothem studies, age modeling considerations will be of secondary
435 importance: in studies of glacial-interglacial cycles, for instance, age offsets of
436 a few decades are immaterial. However, the interpretation of $\delta^{18}\text{O}$ in climate
437 proxies (whether in terms of rainfall, temperature, or other factors) is usually
438 complex, and therefore should be backed by isotope-enabled models (such as
439 LeGrande and Schmidt, 2009; Pausata et al., 2011, and others). We note that
440 proxy system modeling is a burgeoning field with applications to all paleoclimate
441 archives, not just speleothems (e.g. Schmidt, 1999; Evans, 2007; Dee et al., 2015).

442 For the data collecting and measuring stage, high resolution sampling is
443 suggested (von Gunten et al. (2012) suggested collecting 80-100 data points
444 over the calibration period) for time-based calibrations to achieve a sufficient
445 effective sample size. If smoothing scale is known, the record with smaller
446 smoothing scale should be used in the time-based calibration. Also we should

447 note that age uncertainties set the limit of the usage of proxy data. For example,
448 proxy data with decadal age uncertainties cannot be used in interannual-scale
449 research questions, but may be used in centennial-scale research questions (Birks
450 et al., 2012).

451 Our goal in presenting these results is not to indict a particular set of au-
452 thors, as the unsophisticated data-analytical practices of the original studies are
453 unfortunately rather common in the paleoclimate literature. Instead, we wish
454 to draw attention to under-appreciated statistical issues, with the hope of less-
455 ening the occurrences of proxy interpretations based on spurious correlations,
456 and to improve the robustness of future paleoclimate studies. In this spirit,
457 we are making the Python code associated with this study freely available at
458 https://github.com/ClimateTools/Correlation_EPSL in order to dissemi-
459 nate best practices.

460 Acknowledgments

461 We thank Chris Paciorek for his False Discovery Rate code, and thank Andy
462 Baker for providing the weather record of Assynt in Scotland. This work was
463 supported by Grant 1347213 from the US National Science Foundation and by
464 a graduate fellowship from the University of Southern California.

465 References

- 466 Baker, A., Bradley, C., Phipps, S., Fischer, M., Fairchild, I., Fuller, L., Spötl,
467 C., Azcurra, C., 2012. Millennial-length forward models and pseudoproxies
468 of stalagmite $\delta^{18}\text{O}$: an example from NW Scotland. *Climate of the Past* 8,
469 1153–1167. doi:10.5194/cp-8-1153-2012.
- 470 Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A
471 Practical and Powerful Approach to Multiple Testing. *Journal of the Royal*
472 *Statistical Society. Series B (Methodological)* 57, 289–300. doi:10.2307/
473 2346101.

- 474 Birks, J.B., Lotter, A.F., Juggins, S., Smol, J.P., 2012. Tracking environmen-
475 tal change using lake sediments: data handling and numerical techniques.
476 volume 5. Springer Netherlands. doi:10.1007/978-94-007-2745-8.
- 477 Blaauw, M., Christen, J.A., 2011. Flexible paleoclimate age-depth models using
478 an autoregressive gamma process. *Bayesian Analysis* 6, 457–474. doi:10.
479 1214/ba/1339616472.
- 480 Black, B.A., Griffin, D., van der Sleen, P., Wanamaker, A.D., Speer, J.H.,
481 Frank, D.C., Stahle, D.W., Pederson, N., Copenheaver, C.A., Trouet, V.,
482 et al., 2016. The value of crossdating to retain high-frequency variability,
483 climate signals, and extreme events in environmental proxies. *Global Change*
484 *Biology* doi:10.1111/gcb.13256.
- 485 Buening, N.H., Stott, L., Kanner, L., Yoshimura, K., 2013. Diagnosing at-
486 mospheric influences on the interannual $^{18}\text{O}/^{16}\text{O}$ variations in Western US
487 precipitation. *Water* 5, 1116–1140. doi:10.3390/w5031116.
- 488 Cheng, H., Edwards, R.L., Shen, C.C., Polyak, V.J., Asmerom, Y., Woodhead,
489 J., Hellstrom, J., Wang, Y., Kong, X., Spötl, C., et al., 2013. Improvements
490 in ^{230}Th dating, ^{230}Th and ^{234}U half-life values, and U–Th isotopic mea-
491 surements by multi-collector inductively coupled plasma mass spectrometry.
492 *Earth and Planetary Science Letters* 371, 82–91. doi:10.1016/j.epsl.2013.
493 04.006.
- 494 Clark, R., Thompson, R., 1979. A new statistical approach to the alignment of
495 time series. *Geophysical Journal International* 58, 593–607. doi:10.1111/j.
496 1365-246X.1979.tb04796.x.
- 497 Crowley, T.J., 1999. Correlating high-frequency climate variations. *Paleo-*
498 *ceanography* 14, 271–272. doi:10.1029/1999PA900003.
- 499 Dansgaard, W., 1964. Stable isotopes in precipitation. *Tellus* 16, 436–468.
500 doi:10.1111/j.2153-3490.1964.tb00181.x.

501 Dawdy, D., Matalas, N., 1964. Statistical and probability analysis of hydrologic
502 data, part III: Analysis of variance, covariance and time series. McGraw-Hill.

503 Dee, S., Emile-Geay, J., Evans, M.N., Allam, A., Steig, E.J., Thompson, D.M.,
504 2015. PRYSM: An open-source framework for PRoxY System Modeling, with
505 applications to oxygen-isotope systems. *Journal of Advances in Modeling*
506 *Earth Systems* 7, 1220–1247. doi:10.1002/2015MS000447.

507 Emile-Geay, J., 2016. *Data Analysis in the Earth & Environmental Sciences*.
508 Second ed., FigShare. doi:10.6084/m9.figshare.1014336.

509 Emile-Geay, J., Tingley, M., 2016. Inferring climate variability from nonlinear
510 proxies: application to palaeo-ENSO studies. *Climate of the Past* 12, 31–50.
511 doi:10.5194/cp-12-31-2016.

512 Evans, M.N., 2007. Toward forward modeling for paleoclimatic proxy signal
513 calibration: A case study with oxygen isotopic composition of tropical woods.
514 *Geochem. Geophys. Geosyst* 8, Q07008. doi:10.1029/2006GC001406.

515 Ghil, M., Allen, M., Dettinger, M., Ide, K., Kondrashov, D., Mann, M.,
516 Robertson, A.W., Saunders, A., Tian, Y., Varadi, F., et al., 2002. Ad-
517 vanced spectral methods for climatic time series. *Reviews of Geophysics* 40.
518 doi:10.1029/2001RG000092.

519 von Gunten, L., Grosjean, M., Kamenik, C., Fujak, M., Urrutia, R., 2012.
520 Calibrating biogeochemical and physical climate proxies from non-varved lake
521 sediments with meteorological data: methods and case studies. *Journal of*
522 *Paleolimnology* 47, 583–600. doi:10.1007/s10933-012-9582-9.

523 Haam, E., Huybers, P., 2010. A test for the presence of covariance between
524 time-uncertain series of data with application to the Dongge Cave speleothem
525 and atmospheric radiocarbon records. *Paleoceanography* 25. doi:10.1029/
526 2008PA001713.

527 Haslett, J., Parnell, A., 2008. A simple monotone process with application to
528 radiocarbon-dated depth chronologies. *J. R. Stat. Soc. Ser. C-Appl. Stat.* 57,
529 399–418. doi:10.1111/j.1467-9876.2008.00623.x.

530 Ioannidis, J., 2005. Why most published research findings are false. *PLoS Med*
531 2, e124. doi:10.1371/journal.pmed.0020124.

532 Jones, P., Briffa, K., Osborn, T., Lough, J., van Ommen, T., Vinther, B.,
533 Luterbacher, J., Wahl, E., Zwiers, F., Mann, M., Schmidt, G., Ammann, C.,
534 Buckley, B., Cobb, K., Esper, J., Goosse, H., Graham, N., Jansen, E., Kiefer,
535 T., Kull, C., Kuttel, M., Mosley-Thompson, E., Overpeck, J., Riedwyl, N.,
536 Schulz, M., Tudhope, A., Villalba, R., Wanner, H., Wolff, E., Xoplaki, E.,
537 2009. High-resolution palaeoclimatology of the last millennium: a review of
538 current status and future prospects. *The Holocene* 19, 3–49. doi:10.1177/
539 0959683608098952.

540 Kaplan, A., Cane, M.A., Kushnir, Y., Clement, A.C., Blumenthal, M.B., Ra-
541 jagopalan, B., 1998. Analyses of global sea surface temperature 1856-1991. *J.*
542 *Geophys. Res.-Oceans* 103, 18567–18589. doi:10.1029/97JC01736.

543 Khider, D., Huerta, G., Jackson, C., Stott, L., Emile-Geay, J., 2015. A Bayesian,
544 multivariate calibration for *Globigerinoides ruber* Mg/Ca. *Geochemistry,*
545 *Geophysics, Geosystems* 16, 2916–2932. doi:10.1002/2015GC005844.

546 LeGrande, A., Schmidt, G., 2009. Sources of Holocene variability of oxygen
547 isotopes in paleoclimate archives. *Clim. Past* 5, 441–455. doi:10.5194/
548 cp-5-441-2009.

549 Macias-Fauria, M., Grinsted, A., Helama, S., Holopainen, J., 2012. Persistence
550 matters: estimation of the statistical significance of paleoclimatic reconstruc-
551 tion statistics from autocorrelated time series. *Dendrochronologia* 30, 179–
552 187. doi:10.1016/j.dendro.2011.08.003.

553 McCabe-Glynn, S., Johnson, K.R., Strong, C., Berkelhammer, M., Sinha, A.,
554 Cheng, H., Edwards, R.L., 2013. Variable North Pacific influence on drought

555 in southwestern North America since AD 854. *Nat. Geosci.* 6, 617–621. doi:10.
556 1038/NGE01862.

557 Mudelsee, M., 2001. The phase relations among atmospheric CO₂ content,
558 temperature and global ice volume over the past 420 ka. *Quaternary Science*
559 *Reviews* 20, 583–589. doi:10.1016/S0277-3791(00)00167-0.

560 Mudelsee, M., 2013. *Climate time series analysis*. Springer. doi:10.1007/
561 978-90-481-9482-7.

562 Oster, J.L., Montañez, I.P., Santare, L.R., Sharp, W.D., Wong, C., Cooper,
563 K.M., 2015. Stalagmite records of hydroclimate in central California during
564 termination 1. *Quaternary Science Reviews* 127, 199 – 214. doi:10.1016/j.
565 quascirev.2015.07.027.

566 Parnell, A.C., Buck, C.E., Doan, T.K., 2011. A review of statistical chronol-
567 ogy models for high-resolution, proxy-based holocene palaeoenvironmental
568 reconstruction. *Quaternary Science Reviews* 30, 2948–2960. doi:10.1016/j.
569 quascirev.2011.07.024.

570 Partin, J., Quinn, T., Shen, C.C., Emile-Geay, J., Taylor, F., Maupin, C., Lin,
571 K., Jackson, C., Banner, J., Sinclair, D., Huh, C.A., 2013. Multidecadal rain-
572 fall variability in South Pacific Convergence Zone as revealed by stalagmite
573 geochemistry. *Geology* doi:10.1130/G34718.1.

574 Partin, J.W., Jenson, J.W., Banner, J.L., Quinn, T.M., Taylor, F.W., Sin-
575 clair, D., Hardt, B., Lander, M.A., Bell, T., Miklavič, B., et al., 2012. Rela-
576 tionship between modern rainfall variability, cave dripwater, and stalagmite
577 geochemistry in Guam, USA. *Geochemistry, Geophysics, Geosystems* 13.
578 doi:10.1029/2011GC003930.

579 Pausata, F.S., Battisti, D.S., Nisancioglu, K.H., Bitz, C.M., 2011. Chinese sta-
580 lagmite $\delta^{18}\text{O}$ controlled by changes in the Indian monsoon during a simulated
581 Heinrich event. *Nature Geoscience* 4, 474–480. doi:10.1038/ngeo1169.

582 Proctor, C., Baker, A., Barnes, W., Gilmour, M., 2000. A thousand year
583 speleothem proxy record of North Atlantic climate from Scotland. *Climate*
584 *Dynamics* 16, 815–820. doi:10.1007/s003820000077.

585 Rehfeld, K., Kurths, J., 2014. Similarity estimators for irregular and age-
586 uncertain time series. *Climate of the Past* 10, 107–122. doi:10.5194/
587 cp-10-107-2014.

588 Schmidt, G.A., 1999. Forward modeling of carbonate proxy data from plank-
589 tonic foraminifera using oxygen isotope tracers in a global ocean model. *Pa-*
590 *leoceanography* 14, 482–497. doi:10.1029/1999PA900025.

591 Scholz, D., Hoffmann, D.L., 2011. StalAge—An algorithm designed for con-
592 struction of speleothem age models. *Quaternary Geochronology* 6, 369–382.
593 doi:10.1016/j.quageo.2011.02.002.

594 Scholz, D., Hoffmann, D.L., Hellstrom, J., Ramsey, C.B., 2012. A comparison of
595 different methods for speleothem age modelling. *Quaternary Geochronology*
596 14, 94–104. doi:10.1016/j.quageo.2012.03.015.

597 Schulz, M., Mudelsee, M., 2002. REDFIT: estimating red-noise spectra directly
598 from unevenly spaced paleoclimatic time series. *Computers & Geosciences*
599 28, 421–426.

600 Spötl, C., Fairchild, I.J., Tooth, A.F., 2005. Cave air control on dripwater
601 geochemistry, Obir Caves (Austria): Implications for speleothem deposition
602 in dynamically ventilated caves. *Geochimica et Cosmochimica Acta* 69, 2451–
603 2468. doi:10.1016/j.gca.2004.12.009.

604 Storey, J.D., 2002. A direct approach to false discovery rates. *Journal of the*
605 *Royal Statistical Society: Series B (Statistical Methodology)* 64, 479–498.
606 doi:10.1111/1467-9868.00346.

607 Thomson, D.J., 1982. Spectrum estimation and harmonic-analysis. *Proc. IEEE*
608 70, 1055–1096. doi:10.1109/PROC.1982.12433.

- 609 Tierney, J.E., Tingley, M.P., 2014. A Bayesian, spatially-varying calibration
610 model for the TEX86 proxy. *Geochimica et Cosmochimica Acta* 127, 83–106.
611 doi:10.1016/j.gca.2013.11.026.
- 612 Tingley, M.P., Craigmile, P.F., Haran, M., Li, B., Mannshardt, E., Rajaratnam,
613 B., 2012. Piecing together the past: statistical insights into paleoclimatic
614 reconstructions. *Quaternary Science Reviews* 35, 1 – 22. doi:10.1016/j.
615 quascirev.2012.01.012.
- 616 Trouet, V., Esper, J., Graham, N.E., Baker, A., Scourse, J.D., Frank, D.C.,
617 2009. Persistent positive North Atlantic Oscillation mode dominated the me-
618 dieval climate anomaly. *Science* 324, 78–80. doi:10.1126/science.1166349.
- 619 Ventura, V., Paciorek, C.J., Risbey, J.S., 2004. Controlling the Proportion of
620 Falsely Rejected Hypotheses when Conducting Multiple Tests with Climato-
621 logical Data. *Journal of Climate* 17, 4343–4356. doi:10.1175/3199.1.
- 622 Wilks, D.S., 2011. *Statistical Methods in the Atmospheric Sciences: an Intro-*
623 *duction*. Academic Press, San Diego.
- 624 Wunsch, C., 2003. Greenland—Antarctic phase relations and millennial time-
625 scale climate fluctuations in the Greenland ice-cores. *Quaternary Science*
626 *Reviews* 22, 1631–1646. doi:10.1016/S0277-3791(03)00152-5.
- 627 Yule, G.U., 1926. Why do we sometimes get nonsense-correlations between
628 time-series – a study in sampling and the nature of time-series. *Journal of*
629 *the Royal Statistical Society* 89, 1–63. doi:10.2307/2341482.
- 630 Zhu, M., Stott, L., Buckley, B., Yoshimura, K., Ra, K., 2012. Indo-Pacific
631 Warm Pool convection and ENSO since 1867 derived from Cambodian pine
632 tree cellulose oxygen isotopes. *Journal of Geophysical Research: Atmospheres*
633 117. doi:10.1029/2011JD017198.