Putting the significance of spectral peaks on the level: implications for the 1470-yr peak in Greenland δ18O

Peter Huybers[1]

[1]Harvard University

November 26, 2022

Abstract

Spectral analyses of the Greenland Ice Sheet Project 2 (GISP2) δ18O record has been interpreted to show a 1/(1470 yr) spectral peak that is highly statistically significant ($p<0.01$) relative to a null model, H0, consisting of an auto-regressive order one process. In this study H0 is generalized using an auto-regressive moving-average process that is removed from the GISP2 δ18O time series, yielding a spectral estimate that is approximately level. A rule of thumb is proposed for evaluating the adequacy of H0 involving comparing the expected and observed variance of the logarithm of a leveled spectral estimate. After suitably leveling GISP2 δ18O and accounting for multiple hypothesis testing, multitaper spectral estimation indicates that the 1/(1470 yr) peak is insignificant ($p>0.05$). The proposed techniques for fitting H0 should be generally applicable to evaluating peaks in other geophysical spectral estimates.

Putting the significance of spectral peaks on the level: implications for the 1470-yr peak in Greenland δ^{18}O

Peter Huybers

Harvard University

Key Points:

- A rule of thumb is proposed to select a null hypothesis for testing spectral peaks.
- Application to GISP2 δ^{18}O indicates that an auto-regressive order one process is insufficient.
- An adequate null hypothesis indicates that the 1/(1470 year) peak observed in GISP2 δ^{18}O is insignificant.

Corresponding author: P. Huybers, phuybers@fas.harvard.edu

Abstract

Spectral analyses of the Greenland Ice Sheet Project 2 (GISP2) δ^{18}O record has been interpreted to show a 1/(1470 yr) spectral peak that is highly statistically significant (p <0.01) relative to a null model, H_0, consisting of an auto-regressive order one process. In this study H_0 is generalized using an auto-regressive moving-average process that is removed from the GISP2 δ^{18}O time series, yielding a spectral estimate that is approximately level. A rule of thumb is proposed for evaluating the adequacy of H_0 involving comparing the expected and observed variance of the logarithm of a leveled spectral estimate. After suitably leveling GISP2 δ^{18}O and accounting for multiple hypothesis testing, multitaper spectral estimation indicates that the 1/(1470 yr) peak is insignificant (p >0.05). The proposed techniques for fitting H_0 should be generally applicable to evaluating peaks in other geophysical spectral estimates.

Plain Language Summary

A suitable null hypothesis is necessary for obtaining accurate test results, but a means for evaluating the adequacy of a test for a spectral peak has been lacking. A generalized null model is presented in the form of an auto-regressive, moving-average model whose adequacy can be gauged by comparing the observed and expected variance of log spectral coefficients. Application of the method to the GISP2 δ^{18}O record indicates that a much discussed spectral peak found at 1/(1470 year) is statistically insignificant. The seeming prominence of the 1/(1470 year) peak is explained as the result of evaluating a spectrum that is somewhat more complicated than a simple auto-regressive order-one process and the peak having been selected on the basis of its seeming anomalous.

1 Introduction

The mechanisms governing rapid climate change during the last glacial were characterized as unclear more than a decade ago (Clement & Peterson, 2008; Wolff et al., 2010), and this situation is not much changed. A potentially important clue regarding the origin of abrupt climate change involves indications of underlying quasi-periodic behavior. A number of studies pointed to an approximate 1500 year cycle associated with rapid warmings known as Dansgaard-Oeschger (DO) events (Bond et al., 1997; Alley et al., 2001; Schulz, 2002; Rahmstorf, 2003). Postulated models involving periodicity include stochastic resonance (Alley et al., 2001), coherent resonance (Timmermann et al., 2003),

or nonlinear pacing of DO events (Schulz, 2002; Rahmstorf, 2003; Braun et al., 2005). The distribution of waiting times between DO events was later argued, however, to also be consistent with an exponential distribution (Ditlevsen et al., 2005). Whether DO event timing constitutes strong evidence for an underlying quasi-periodic process is unresolved (Wolff et al., 2010).

A second line of evidence for underlying periodicity in millennial glacial climate variability comes from spectral analysis. δ^{18}O variations measured in the Greenland Ice sheet Project 2 (GISP2) record from the last glacial show a strong spectral peak at frequencies near 1/(1470 yr) (Yiou et al., 1997). A similar peak was also identified in other GISP2 ice-core measurements (Mayewski et al., 1997) and marine sediment core records of similar age from the North Atlantic (Stocker & Mysak, 1992). Previous analyses suggest that the 1/(1470 yr) peak in Greenland δ^{18}O is highly statistically significant ($p < 0.01$) on the basis of comparing the spectral estimate against a null model, H_0, consisting of an auto-regressive order one process, AR(1) (Schulz, 2002; Schulz & Mudelsee, 2002). It is known that DO timing is related to the spectral peak at 1/(1470 years) (Schulz, 2002), but absent an accepted model for the timing of DO events, the exact relationship is unclear.

As emphasized some time ago (Wunsch, 2000), it would be a remarkable discovery if the DO events, or other aspects of last glacial climate, exhibited quasi-periodic behavior. Such behavior would indicate the presence of important quasi-periodic forcing, internal processes that cause the climate system to be highly sensitive in a narrow band of frequencies, or some combination of both. Because such a discovery would be remarkable, it should be scrutinized.

In the following, the significance of 1/(1470 yr) spectral peak found in GISP2 δ^{18}O is examined. Focus is upon selecting an appropriate H_0 because, as pointed out generally by Vaughan et al. (2011), insufficient representations of H_0 can readily lead to assigning spurious significance to spectral peaks. Spectral estimates of geophysical time series are commonly evaluated relative to a null spectral model based on an AR(1) process (e.g. Mann & Lees, 1996; Von Storch, 1999). Ditlevsen et al. (2005), however, pointed out that an AR(1) process gives a poor fit to the GISP2 δ^{18}O spectral estimate. Other processes have also been proposed for representing H_0 including integrated and moving-averages (Box et al., 2015; Klaus et al., 2015), use of a portion of the sample auto-correlation

function (Priestley, 1981; Garrido & García, 1992; Goff, 2020), and power-law distributions (Vaughan, 2005). Which of these processes, if any, is an adequate representation of H_0 for GISP2 δ^{18}O during the last glacial is unclear.

2 Methods

2.1 A null spectral model

A model of H_0 is postulated that admits for fitting the observed spectral estimate to varying detail. By Wold's theorom, either an auto-regressive (AR) or moving average (MA) model can represent any stationary time-series given a sufficiently high order. Whether GISP2 δ^{18}O is stationary is highly questionable, but it is nevertheless found that using a combined AR and MA process allows for a parsimonious representation of H_0 for the GISP2 δ^{18}O spectral estimate. The ARMA model is,

$$x(t) = \epsilon_t + \sum_{n=1}^{p} a_n x_{t-n} + \sum_{n=1}^{q} m_n \epsilon(t-n), \tag{1}$$

where $x(t)$ is the time-series of interest, a_n are AR coefficients up to order p, and m_n are MA coefficients up to order q (Box et al., 2015). The terms involving $\epsilon(t)$ represent the innovations, or, new information at each time step, that the ARMA process acts upon. The various null models that will be considered are expressed in terms of the order of the AR and MA process involved, $H_0(p,q)$.

Eq. 1 is fit in the time domain using an iterative maximum likelihood approach (Box et al., 2015). For purposes of illustrating H_0, note that the spectral representation of the associated ARMA process is,

$$H_0^*(s) = \left| \frac{1 + \sum_{n=1}^{q} m_n e^{-i2\pi s n \delta t}}{1 - \sum_{n=1}^{p} a_n e^{-i2\pi s n \delta t}} \right|^2, \tag{2}$$

where δt is the time interval between observations in x, and s is frequency.

The multitaper method (Thomson, 1982) is used to estimate spectral density, $P(s)$.

If $P(s)$ is estimated using $\epsilon(t)$, as opposed to $x(t)$, the result is referred to as having been leveled according to $H_0(p,q)$. In particular, if $\epsilon(t)$ is independent and normally distributed, the expected value of P is level in the sense of being constant with frequency, an operation also referred to as pre-whitening (Priestley, 1981). Values for $\epsilon(t)$ are obtained by inverting $x(t)$ conditional on the specified ARMA coefficients (Box et al., 2015).

If level, the distribution of P is expected to follow a gamma distribution,

$$f(P) = \frac{P^{k-1}}{\Gamma(k)\theta^k} e^{-\frac{P}{\theta}}, \tag{3}$$

where Γ is the gamma function. The shape parameter, k, equals the degrees of freedom in the spectral estimate divided by two. For a multitaper spectral estimate, k equals the number of tapers applied. The scale parameter, θ, equals σ^2/k, where σ^2 is the variance of $\epsilon(t)$ when $P(s)$ is normalized such that its mean across frequency equals the variance of $\epsilon(t)$. Note that setting $\theta = 2$ and substituting $k = d/2$ in Eq. 3 gives the chi-square distribution with d degrees of freedom.

Eq. 3 can be transformed using the natural logarithm of P,

$$f(\ln P) = \frac{P^k}{\Gamma(k)\theta^k} e^{-\frac{P}{\theta}}, \tag{4}$$

for which there are two principal advantages. First, the variance of $\ln P$ only depends on the degrees of freedom in the estimate, $\Psi_3(k)$, where Ψ_3 is the trigamma function. In order to gauge the adequacy of H_0, Ψ_3 is compared against the the sample variance of the logarithm of the leveled spectral estimate,

$$\psi = \frac{N}{2} \sum_s (\ln P(s))^2. \tag{5}$$

N is the number of data points in $\epsilon(t)$ and is assumed even.

A second advantage to using the logarithmic transform is that the variance of $\ln P$ is less sensitive to concentrations of spectral energy than P. In Section 4 it is shown that ψ is not especially biased by a quasi-periodic contribution consistent with the GISP2 observations, though a larger quasi-periodic contribution could bias ψ high to the point of indicating an overly-stringent formulation of H_0. In such a case, an iterative approach can be applied whereby the most significant spectral peak is removed, H_0 refit, and the process repeated until no further significant peaks are identified (Hannan, 1961).

Eq. 5 provides a rule-of-thumb for purposes of evaluating whether H_0 is plausible. More-specific tests for the goodness-of-fit of $\ln P$ to a log-gamma distribution could be applied such as the Kolmogorov–Smirnov or Cramér–von Mises tests. The significance of such a test, however, would need to be interpreted in the context of the autocorrelation between neighboring spectral density estimates and demand a Monte Carlo evaluation of significance levels. Section 4 presents Monte Carlo realizations but that are focused on evaluating the significance of the spectral peak directly. Box and Pierce (1970)

also gave a time-domain test for whether there remains significant residual autocorrelation in $\epsilon(t)$ after applying an ARMA model but which is sensitive to the presence of a quasi-periodic component.

Two other features of $\ln P$ are also worth noting. First, the mean of $\ln P$ is biased negative by $b = \Psi_2(k) + \ln 2 - \ln 2k$, which becomes relevant when computing confidence levels. Ψ_2 is the digamma function. Second, $\ln P$ is more normally distributed than P. Whereas the skewness of P is $\sqrt{8/k}$, it is $\Psi_4(k)/(\Psi_3(k))^{3/2}$ for $\ln P$ (Olshen, 1938). Ψ_4 is the tetragamma function. It follows that $\ln P$ has a skewness that is 0.57 times that of P for $k = 2$, with the fraction asymptoting to 0.5 as k becomes large.

2.2 Multiple hypothesis testing

The fact that the 1/(1470 yr) peak in GISP2 δ^{18}O was selected on the basis of its appearing especially large (Yiou et al., 1997) implies that if another frequency had shown a large peak, it too would have been considered for significance, necessitating accounting for a multiple hypothesis testing regime. Thomson (1990) suggested using a significance level of $(1 - 1/N)$ when evaluating the full spectral estimate for periodic components, where N is the number of frequencies in a spectral estimate. In evaluating GISP2 δ^{18}O, Schulz and Mudelsee (2002) applied Thomson's rule-of-thumb to Welch's method of spectral estimation. Although preferable to no correction, Thomson's suggestion does not account for the degrees of freedom in a spectral estimate, and a more detailed accounting appears useful.

Multiple hypothesis testing can be accounted for using a Bonferroni correction (e.g. Vaughan et al., 2011) whereby the significance level is divided by the number of independent tests. The number of individual tests in a multitaper spectral estimate is estimated as the ratio of the frequency range to the bandwidth resolution of a multitaper spectral estimate. Frequency ranges from zero to the Nyquist frequency, $1/(2\Delta t)$, and the bandwidth resolution of a multitaper spectral estimate is $k/(N\Delta t)$, where k is the number of tapers and N is the number of data points in the time series. The number of independent tests is, thus, $N/(2k)$ and the Bonferroni-adjusted significance level is,

$$\alpha_b = 1 - \alpha 2k/N. \tag{6}$$

Following previous analyses (Schulz, 2002; Schulz & Mudelsee, 2002), an $\alpha = 0.01$ significance level is specified. For purposes of illustrating the sensitivity of results to different test protocols, a significance level of $\alpha = 0.05$ is also discussed.

Eq. 6 requires no assumptions regarding dependencies between tests. An assumption that tests are not negatively correlated would also be appropriate and allows for a slightly more powerful test (Šidák, 1967) but would not change any of the results presented herein. Differences in accounting for multiple hypothesis testing may become relevant, however, in cases involving a very large number of independent frequency bands or tests using larger α values.

3 Analysis of GISP2 δ^{18}O

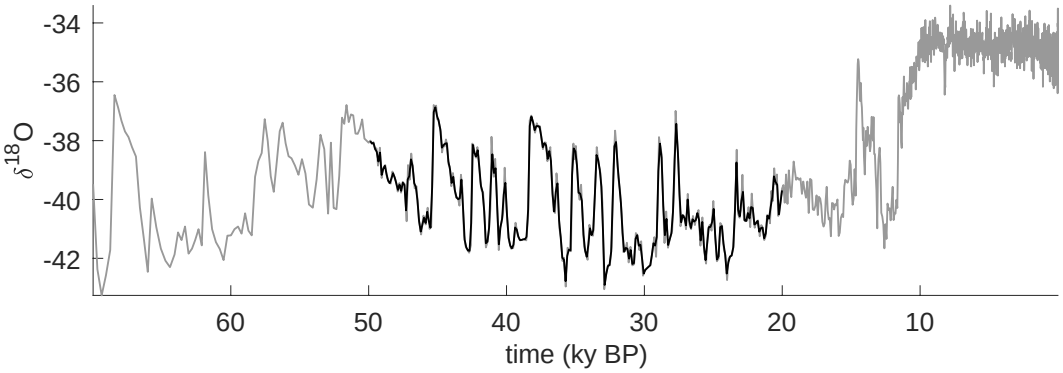

Figure 1. Greenland Ice Sheet Project 2 $\delta^{18}O$ observations. Shown are both the original record at native sampling resolution (gray) and a 50 to 20 ka interval that is interpolated to uniform resolution (black).

Figure 1 shows the δ^{18}O record from the Greenland Ice Sheet Project 2 (GISP2) with ages from stratigraphic layer counting (Meese et al., 1997). Focus is on the 20 to 50 ka interval as representing a relatively homogeneous interval of the last glacial. δ^{18}O samples have an average spacing of $\Delta t = 124$ yr between 20 to 50 ka, with the time between neighboring pairs of data points ranging between 77 to 306 yr. To place the record on an evenly-sampled age scale, it is first linearly interpolated to 12.4 yr resolution, then smoothed to remove variability at periods shorter than 124 yr by convolving with an 11-point Hamming window, and, finally, decimated to 124-yr resolution in order to maintain the same average sampling interval. Interpolation to high-resolution followed by smooth-

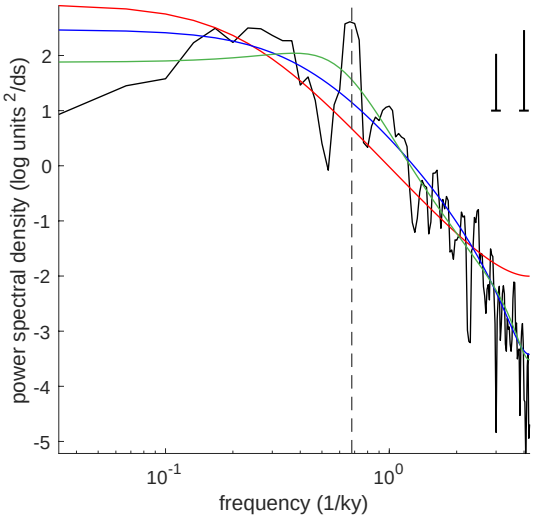

Figure 2. Multitaper spectral estimate of the GISP2 δ^{18}O record between 50 to 20 ka. The spectral estimate is made using five tapers (black line, 10 degrees of freedom) and units are the natural logarithm of spectral power density. Candidate null spectral models, H_0, are fit using auto-regressive processes of order p and moving-average processes of order q, or $H_0(p,q)$ processes. Increasingly detailed versions of H_0 indicate that the spectral peak at 1/(1470 yr) (vertical dashed line) is less pronounced: H_0(1,0) in red, H_0(1,1) in blue, and H_0(3,2) in green. The distance above H_0 corresponding to the 99% confidence level is shown assuming a single test (left vertical black line, shorter) and a multiple hypothesis test (right, using Eq. 6).

ing and decimation helps reduce aliasing and makes results less sensitive to the exact specification of the interpolation grid. Techniques are available for computing spectral estimates directly from unevenly-spaced data, thereby circumventing biases resulting from interpolation, but such estimates are biased for other reasons and require further correction (Schulz & Mudelsee, 2002).

Applying multitaper spectral estimation with three tapers, $k = 3$, gives a spectral estimate having the well-noted (Yiou et al., 1997; Wunsch, 2000; Schulz & Mudelsee, 2002; Ditlevsen et al., 2005) spectral peak centered at 1/(1470 yr) (Figure 2). Although dependencies in time series can reduce the effective degrees of freedom in a spectral estimate, this issue is almost entirely circumvented if the time series is first leveled. For example, the effective number of degrees of freedom at frequencies above 1 cycle per ky in the GISP2 δ^{18}O record averages 5.3 according to the estimation algorithm given by Percival et al. (1993), but the effective degrees of freedom is 5.95 after leveling accord-

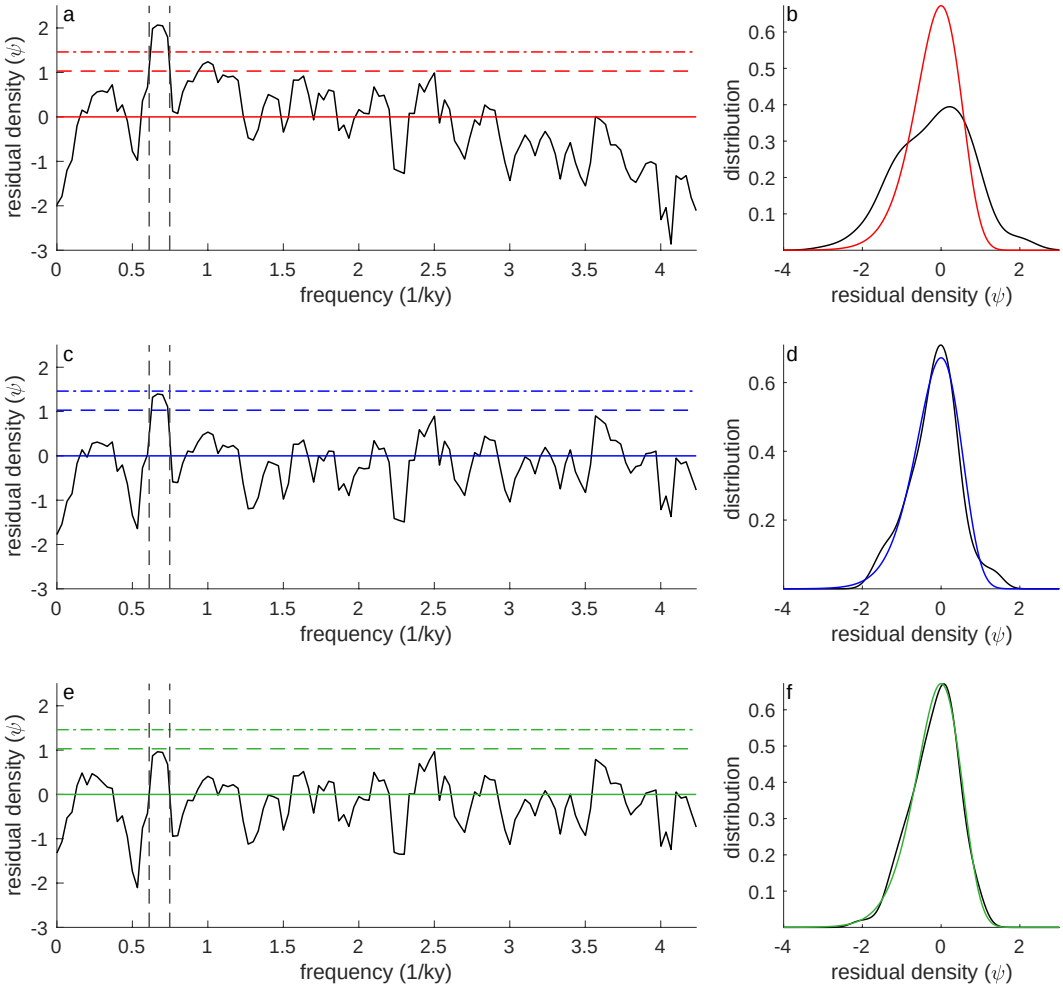

Figure 3. Leveled spectral estimates show differing significance levels. The GISP2 δ^{18}O record shown in Fig. 1 is leveled using the three version of H_0 depicted in Fig. 2. Leveled spectral density is expected to follow a log-gamma distribution (Eq. 2) with variance $\Psi_3(k = 3) = 0.39$ (Eq. 3). (a) Leveling according to H_0(1,0) indicates that the 1/(1470 yr) is significant at the $\alpha = 0.01$ (dashed-dot line) confidence level, but (b) the sample variance is $\psi = 0.87$. (c,d) Similar to a,b but where H_0(1,1) indicates the 1/(1470 yr) peak is just significant relative to the $\alpha = 0.05$ confidence level (dashed line), and $\psi = 0.38$. (e,f) Under H_0(3,2) the 1/(1470 yr) peak is insignificant, and $\psi = 0.35$.

ing to H_0(1,0) and even closer to 6 if using a higher-order representation of H_0. Degrees of freedom are, thus, adequately approximated as equalling two times k in leveled time series.

All combinations of H_0 involving p from 0 to 5 and q from 0 to 5 are considered, but for specificity three cases are focused upon: $H_0(1,0)$ has been used previously, $H_0(1,1)$ is an edge case that is significant at $\alpha = 0.01$ but not $\alpha = 0.05$, and results from higher-order processes are illustrated by $H_0(3,2)$. Although ARMA models are fit in the time domain and time series leveled prior to spectral analysis, it is useful to plot the spectral representation of the ARMA process (Eq. 2). Each version of H_0 considered here gives a smooth fit to the spectral estimate, but with the implied height of the 1/(1470 yr) peak above H_0 diminishing with higher-order ARMA processes (Figure 2).

The confidence level of a properly leveled spectral estimate is constant. For a single hypothesis test with $k = 3$ and $\alpha = 0.01$ the confidence level is 1.03 above the mean, and for a multiple hypothesis version (Eq. 6) with $N = 297$ data points it is 1.46. For $\alpha = 0.05$ the single and multiple hypothesis test confidence levels are, respectively, 0.74 and 1.30. Units are suppressed because $\ln P$ is normalized by removing its sample mean. Removing the sample mean of $\ln P$ is equivalent to dividing P by its sample mean except that mean of the logarithm must be corrected for bias, such that spectral results are expressed as $\ln P - (\overline{\ln P} - b)$, where b is given in Section 2. The normalized value of $\ln P$ is expected to follow a log-gamma distribution centered on zero insomuch as H_0 is adequate. Normalization slightly reduces the degrees of freedom and increases the significance level, but this change is minor relative to differences among versions of H_0.

The spectral estimate obtained after leveling under H_0(1,0) indicates that the 1/(1470 yr) spectral peak is significant at $\alpha = 0.01$ (Fig. 3a). The spectral estimate is not level, however, having a clear trend toward lower spectral density between millennial and shorter-period variations. This systematic structure calls into question the validity of confidence levels computed under the log-gamma assumption and is reflected in the fact that the variance of the leveled spectral estimate, $\psi = 0.80$, is more than twice that expected, $\Psi_3 = 0.39$ (Fig. 3b).

Leveling according to H_0(1,1) gives a spectral estimate whose variance is close to that expected, $\psi = 0.38$, and the spectral peak at 1/(1470 yr) is then insignificant at the $\alpha = 0.01$ level. Under H_0(3,2) the residual variance is $\psi = 0.35$ and the 1/(1470 yr) is insignificant even at $\alpha = 0.05$. More generally, values of ψ are within 13% of Ψ_3 for all examined version of H_0 with the notable exception of H_0(1,0), which is 220% larger, and minor exception of H_0(5,5), which is 16% lower. Only H_0(1,0) indicates that the 1/(1470

yr) peak is significant at $\alpha = 0.01$, and only $H_0(1,0)$, $H_0(1,1)$, and $H_0(1,2)$ indicate significance at $\alpha = 0.05$.

4 Simulated results

Simulations are useful for purposes of better distinguishing among the different versions of H_0. $H_0(1,1)$ is simulated by realizing time-series according to the ARMA(1,1) coefficients fit to the GISP2 record. H_1 is equivalently simulated but also contains a sinusoidal contribution at the 1/(1470 yr) frequency. Innovations in the ARMA process are assigned unit variance and, for H_1, the amplitude of the sinusoidal component is specified to equal 1.125 such that, on average, 23% of the spectral energy resides within $1/1470 \pm 0.1/1470$ yr^{-1}, consistent with observational estimates (Figure 2). 10^4 realizations of H_0 and H_1 are each evaluated using null models that are under specified, perfect, and, over-specified. For each trial, the spectral estimate at the 1/(1470 yr) frequency and the maximum spectral estimate across frequency are recorded, where the latter is the focus on account of being in a multiple hypothesis testing regime.

Leveling the simulated data using an under-specified model, $H_0(1,0)$, leads to a false rejection rate of H_0 of 57% when assessing the significance of the maximum spectral estimate (Fig. 4a). Given that the target false rejection rate is 1%, $H_0(1,0)$ is deeply flawed. Using the simulated $\alpha = 0.01$ significance level under H_0, as opposed to the value obtained assuming a log-gamma distribution, indicates that the 1/(1470 yr) peak is not significant. $H_0(1,0)$ results in a large residual spectral variance with an average ψ of 0.85, compared to an expected value of $\Psi_3 = 0.39$ (Fig. 4b).

Use of a perfect model, $H_0(1,1)$, results in a false rejection rate of 1.0%, as intended (Fig. 4c). These results support the accuracy of the multiple hypothesis testing correction. The residual variance is similar to the expected values, averaging $\psi = 0.41$ (Fig. 4d). $H_0(1,1)$ results support the utility of the variance metric, ψ, in terms of its being discriminating and not strongly biased by the presence of a periodic component. The 99th percentile of ψ under $H_0(1,1)$ is 0.72, a value that 74% of the ψ realizations under $H_0(1,0)$ exceeds, indicating that deficiencies of a magnitude similar to those associated with $H_0(1,0)$ should be routinely identifiable. Furthermore, the mean of ψ under $H_1(1,1)$ is 0.46, or only 13% larger than under H_0, indicating that the variance metric is not overly sensi-

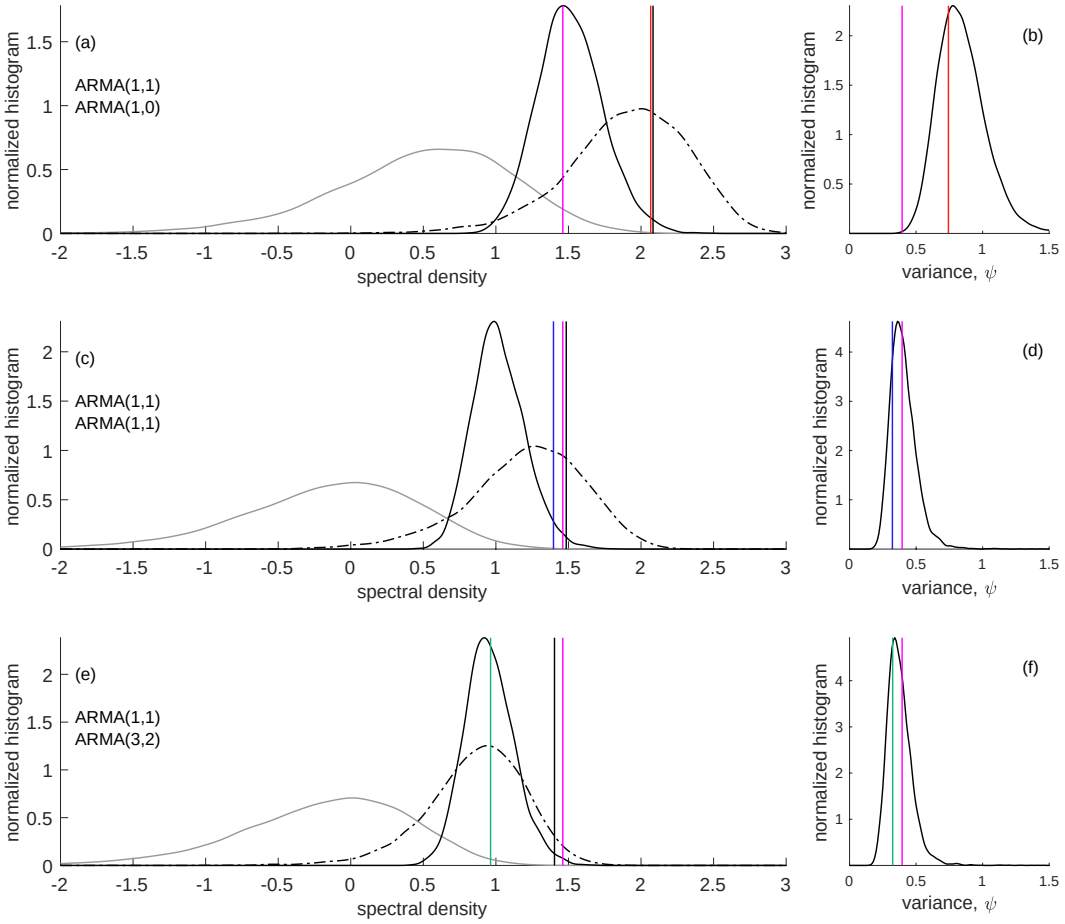

Figure 4. Spectral analysis of synthetic data using different ARMA leveling processes. Simulated data is realized according to an ARMA(1,1) process fit to the GISP2 record and analysed using the multitaper method. Time-series are variously leveled using (a,b) H_0(1,0), (b,c) H_0(1,1), or (d,e) H_0(3,2) null models. Panels a,c, and d each show two null distributions and one alternate: the distributions of log spectral density at the 1/(1470 yr) frequency (H_0, gray line), the maximum value of each spectral analysis (H_0, black line), and the distribution of 1/(1470 yr) log spectral density from an alternate model including a 1470 yr periodic component (H_1, black dash-dot line). The 99% confidence level for H_0 is shown based on the log-gamma distribution (magenta vertical line) and as obtained from the synthetic realizations (black vertical lines). The observed log spectral energy density is shown using colors that correspond with Figs. 2 and 3 (red, blue, or green). Panels b, d, and f show the distribution of variance of leveled spectra estimates (ψ, black curve), the expected variance (Ψ_1, magenta vertical line), and the observed variance (ψ; red, blue, or green vertical lines).

tive to the presence of a moderate periodic component. If P was evaluated, instead of $\ln P$, spectral estimates would average 70% larger under $H_1(1,1)$ than $H_0(1,1)$.

$H_0(3,2)$ leads to over fitting. One result of over fitting is that the false rejection rate is low at 0.4% (Fig. 4e) and, another, that ψ averages 0.38, or slightly below Ψ_3. Furthermore, $H_0(3,2)$ overlaps with $H_1(3,2)$ such that the test is essentially devoid of statistical power.

These result do not necessarily imply that $H_0(1,1)$ is a better choice than $H_0(3,2)$ for the actual GISP2 δ^{18}O data. A parallel set of synthetic studies shows that if the simulated data is instead made from an ARMA(3,2) process fit to the GISP2 δ^{18}O record, $H_0(1,1)$ is strongly biased towards rejecting the null, with a false rejection rate of 22%, and that $H_0(3,2)$ gives the least-biased results.

5 Further discussion and conclusions

A multitaper spectral estimate of GISP2 δ^{18}O indicates a highly significant ($p < 0.01$) spectral peak at 1/(1470 yr) if using a null model based on an ARMA(1,0) process, $H_0(1,0)$ (Fig. 2,3a). $H_0(1,0)$ is found to be inadequate, however, based upon the variance of the logarithm of the leveled spectral estimate being too large (Fig. 3b) and simulations indicating that the rate of false rejection is greater than 1 in 2 (Fig. 4a,b). Conversely, $H_0(1,1)$ and $H_0(3,2)$ indicate that the leveled spectral variance is consistent with expectations (Fig. 3d,f) and that the spectral peak is insignificant (Fig. 2,3c,e), with $p > 0.01$ for $H_0(1,1)$ and $p > 0.05$ for $H_0(3,2)$.

It appears impossible to rule out the results of the higher-order versions of H_0 on two accounts. First, the residual variance is consistent with the expected variance of the logarithm of the spectral power estimate in simulations using $H_0(1,1)$ or other combinations of higher-order ARMA processes, including $H_0(3,2)$ (Fig. 4). Second, an ARMA process is almost certainly under specified in certain respects given the complex dynamics associated with North Atlantic climate variability and its recording in Greenland δ^{18}O (e.g. Guillevic et al., 2013; Rhines & Huybers, 2014). Absent a convincing physical model for climate forcing at a period near 1470 years or a climate system that would respond sensitively in a narrow band near 1/(1470 years), it appears best to regard GISP2 δ^{18}O as entailing a detailed continuum of variability whose description requires an ARMA process higher than AR(1) and probably higher than ARMA(1,1).

A final considerations that vitiates against the 1/(1470 year) peak as being significant is that only selection across frequencies has been considered in the current multiple hypothesis testing regime. The number of hypotheses being tested could, however, be considered larger by some factor. The δ^{18}O record from the Greenland Ice Core Project (GRIP), for example, shows no comparable peak, apparently because of using a flow-based age model as opposed to the layer-counted age model of GISP2 (Ditlevsen et al., 2005). Furthermore, spectral analysis is conducted on the epoch between 20-50 ka, whereas the Holocene does not show variability at 1/(1470 year). Other indicators of North Atlantic variability with approximately a 1500 year period were also found to be confined to a subset of the glacial interval (Obrochta et al., 2012). The number of combinations involving frequency, epoch, and choice of record is, thus, substantially larger than when considering frequency alone, but no specific correction is pursued because the main point is already made.

The seeming prominence of the 1/(1470 year) peak can be understood as relating to its being selected upon the basis of its appearing unusually large, and its being embedded with a continuum of spectral variability whose structure is complex. In this view, rather than being a clue that the underlying mechanism involves quasi-periodicity, the 1/(1470 year) spectral appears a red herring and should be ignored with respect to constraining the processes that govern abrupt climate change.

Acknowledgments

Carl Wunsch (MIT and Harvard) gave helpful feedback on an earlier version of this manuscript. All data and code for computing the results and plotting the figures in this manuscript are temporarily posted at the below link and would be posted to a permanent repository upon publication:

`https://www.dropbox.com/sh/o1bcrfjqey3keoe/AADZyq1Zw7I_Cg39NI8QDbG6a?dl=0`

References

Alley, R., Anandakrishnan, S., & Jung, a. P. (2001). Stochastic resonance in the north atlantic. *Paleoceanography*, *16*(2), 190–198.

Bond, G., Showers, W., Cheseby, M., Lotti, R., Almasi, P., DeMenocal, P., ... Bonani, G. (1997). A pervasive millennial-scale cycle in north atlantic holocene

and glacial climates. *science*, *278*(5341), 1257–1266.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, *65*(332), 1509–1526.

Braun, H., Christl, M., Rahmstorf, S., Ganopolski, A., Mangini, A., Kubatzki, C., ... Kromer, B. (2005). Possible solar origin of the 1,470-year glacial climate cycle demonstrated in a coupled model. *Nature*, *438*(7065), 208–211.

Clement, A. C., & Peterson, L. C. (2008). Mechanisms of abrupt climate change of the last glacial period. *Reviews of Geophysics*, *46*(4).

Ditlevsen, P. D., Kristensen, M. S., & Andersen, K. K. (2005). The recurrence time of dansgaard–oeschger events and limits on the possible periodic component. *Journal of Climate*, *18*(14), 2594–2603.

Garrido, J., & García, J. (1992). Periodic signals in spanish monthly precipitation data. *Theoretical and applied climatology*, *45*(2), 97–106.

Goff, J. A. (2020). "empirical prewhitening" spectral analysis detects periodic but inconsistent signals in abyssal hill morphology at the southern east pacific rise. *Geochemistry, Geophysics, Geosystems*, *21*(11), e2020GC009261.

Guillevic, M., Bazin, L., Landais, A., Kindler, P., Orsi, A., Masson-Delmotte, V., ... others (2013). Spatial gradients of temperature, accumulation and δ 18 o-ice in greenland over a series of dansgaard–oeschger events. *Climate of the Past*, *9*(3), 1029–1051.

Hannan, E. (1961). Testing for a jump in the spectral function. *Journal of the Royal Statistical Society: Series B (Methodological)*, *23*(2), 394–404.

Klaus, J., Chun, K. P., & Stumpp, C. (2015). Temporal trends in δ18o composition of precipitation in germany: insights from time series modelling and trend analysis. *Hydrological Processes*, *29*(12), 2668–2680.

Mann, M. E., & Lees, J. M. (1996). Robust estimation of background noise and signal detection in climatic time series. *Climatic change*, *33*(3), 409–445.

Mayewski, P. A., Meeker, L. D., Twickler, M. S., Whitlow, S., Yang, Q., Lyons, W. B., & Prentice, M. (1997). Major features and forcing of high-latitude northern hemisphere atmospheric circulation using a 110,000-year-long glacio-

chemical series. *Journal of Geophysical Research: Oceans*, *102*(C12), 26345–26366.

Meese, D., Gow, A., Alley, R., Zielinski, G., Grootes, P., Ram, M., ... Bolzan, J. (1997). The greenland ice sheet project 2 depth-age scale: methods and results. *Journal of Geophysical Research: Oceans*, *102*(C12), 26411–26423.

Obrochta, S. P., Miyahara, H., Yokoyama, Y., & Crowley, T. J. (2012). A re-examination of evidence for the north atlantic "1500-year cycle" at site 609. *Quaternary Science Reviews*, *55*, 23–33.

Olshen, A. C. (1938). Transformations of the pearson type iii distribution. *The Annals of Mathematical Statistics*, *9*(3), 176–200.

Percival, D. B., Walden, A. T., et al. (1993). *Spectral analysis for physical applications*. cambridge university press.

Priestley, M. B. (1981). *Spectral analysis and time series: probability and mathematical statistics* (Nos. 04; QA280, P7.).

Rahmstorf, S. (2003). Timing of abrupt climate change: A precise clock. *Geophysical Research Letters*, *30*(10).

Rhines, A., & Huybers, P. J. (2014). Sea ice and dynamical controls on preindustrial and last glacial maximum accumulation in central greenland. *Journal of Climate*, *27*(23), 8902–8917.

Schulz, M. (2002). On the 1470-year pacing of dansgaard-oeschger warm events. *Paleoceanography*, *17*(2), 4–1.

Schulz, M., & Mudelsee, M. (2002). Redfit: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. *Computers & Geosciences*, *28*(3), 421–426.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*(318), 626–633.

Stocker, T. F., & Mysak, L. A. (1992). Climatic fluctuations on the century time scale: A review of high-resolution proxy data and possible mechanisms. *Climatic change*, *20*, 227–250.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, *70*(9), 1055–1096.

Thomson, D. J. (1990). Quadratic-inverse spectrum estimates: applications to

palaeoclimatology. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, *332*(1627), 539–597.

Timmermann, A., Gildor, H., Schulz, M., & Tziperman, E. (2003). Coherent resonant millennial-scale climate oscillations triggered by massive meltwater pulses. *Journal of Climate*, *16*(15), 2569–2585.

Vaughan, S. (2005). A simple test for periodic signals in red noise. *Astronomy & Astrophysics*, *431*(1), 391–403.

Vaughan, S., Bailey, R., & Smith, D. (2011). Detecting cycles in stratigraphic data: Spectral analysis in the presence of red noise. *Paleoceanography*, *26*(4).

Von Storch, H. (1999). Misuses of statistical analysis in climate research. In *Analysis of climate variability* (pp. 11–26). Springer.

Wolff, E. W., Chappellaz, J., Blunier, T., Rasmussen, S. O., & Svensson, A. (2010). Millennial-scale variability during the last glacial: The ice core record. *Quaternary Science Reviews*, *29*(21-22), 2828–2838.

Wunsch, C. (2000). On sharp spectral lines in the climate record and the millennial peak. *Paleoceanography*, *15*(4), 417–424.

Yiou, R., Fuhrer, K., Meeker, L., Jouzel, J., Johnsen, S., & Mayewski, P. A. (1997). Paleoclimatic variability inferred from the spectral analysis of greenland and antarctic ice-core data. *Journal of Geophysical Research: Oceans*, *102*(C12), 26441–26454.