∂Putting the Significance of Spectral Peaks on the Level: Implications for the 1470-Yr Peak in Greenland δ^{18}O

PETER HUYBERS[a]

[a] *Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts*

ABSTRACT: Spectral analysis 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 ($p < 0.01$). The presence of such a peak, if accurate, provides an important clue about the mechanisms controlling glacial climate. As is standard, however, statistical significance was judged relative to a null model, H_0, consisting of an autoregressive order one process, AR(1). In this study, H_0 is generalized using an autoregressive moving-average process, ARMA(p, q). A rule of thumb is proposed for evaluating the adequacy of H_0 involving comparing the expected and observed variances of the logarithm of a spectral estimate, which are generally consistent insomuch as removal of the ARMA structure from a time series results in an approximately level spectral estimate. An AR(1), or ARMA(1, 0), process is shown to be an inadequate representation of the GISP2 δ^{18}O structure, whereas higher-order ARMA processes result in approximately level spectral estimates. 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. The seeming prominence of the 1/(1470 yr) peak is explained as the result of evaluating a spectrum involving higher-order ARMA structure and the peak having been selected on the basis of its seeming anomalous. The proposed technique for evaluating the significance of spectral peaks is also applicable to other geophysical records.

SIGNIFICANCE STATEMENT: A suitable null hypothesis is necessary for obtaining accurate test results, but a means for evaluating the adequacy of a null hypothesis for a spectral peak has been lacking. A generalized null model is presented in the form of an autoregressive, moving-average process whose adequacy can be gauged by comparing the observed and expected variance of log spectral density. Application of the method to the GISP2 δ^{18}O record indicates that spectral structure found at 1/(1470 yr) is statistically insignificant.

KEYWORDS: Climate variability; Paleoclimate; Spectral analysis/models/distribution

1. Introduction

Abrupt warming events recorded in Greenland ice cores during glacial times, known as Dansgaard-Oeschger (DO) events (Dansgaard et al. 1982; Johnsen et al. 1992), are understood to be closely associated with retreat of North Atlantic Ocean sea ice (Li et al. 2005; Sadatzki et al. 2019) and to be part of global variations in climate (Rosen et al. 2014; Corrick et al. 2020). The cause of the DO events remains unclear, however, partly because of the difficulty of resolving a causal sequence in variations that occur over decades (Erhardt et al. 2019).

A potentially important clue about the origin of abrupt climate change involves indications of underlying periodic behavior. A number of studies pointed to an approximate 1500-yr cycle associated with DO events (Bond et al. 1997; Alley et al. 2001; Schulz 2002; Rahmstorf 2003). A related line of evidence comes from spectral analysis of δ^{18}O variations measured in the Greenland Ice sheet Project 2 (GISP2) record from the last glacial that show a spectral peak at

frequencies near 1/(1470 yr) (Yiou et al. 1997). A similar peak was identified in other GISP2 ice-core measurements (Mayewski et al. 1997) and marine sediment core records of similar age from the North Atlantic (Stocker and Mysak 1992). This 1/(1470 yr) peak in Greenland δ^{18}O was argued to be highly statistically significant ($p < 0.01$) on the basis of comparing the spectral estimate with a null model, H_0, consisting of an autoregressive order-one process, AR(1) (Schulz 2002; Schulz and Mudelsee 2002).

The presence of a statistically significant spectral peak at 1/(1470 yr) would provide an important constraint on the mechanisms responsible for DO events (Wunsch 2000; Schulz 2002). Such a peak would indicate the presence of quasi-periodic forcing or internal processes that cause the climate system to be highly sensitive in a narrow band of frequencies. Postulated models for such quasi-periodic variability in glacial climate include stochastic resonance (Alley et al. 2001), coherent resonance (Timmermann et al. 2003), and nonlinear pacing of DO events (Schulz 2002; Rahmstorf 2003; Braun et al. 2005).

The degree to which the timing of DO events is indicative of an underlying quasi-periodic process, however, remains unresolved (Wolff et al. 2010). The distribution of waiting times between DO events was alternatively argued to be consistent with an exponential distribution (Ditlevsen et al. 2007), and the applicability of an AR(1) process for

---

∂Denotes content that is immediately available upon publication as open access.

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

representing H_0 was questioned (Ditlevsen et al. 2005). An insufficient representation of H_0 can lead to assigning spurious significance to spectral peaks (Vaughan et al. 2011). In the following, the significance of the 1/(1470 yr) spectral peak found in GISP2 $\delta^{18}O$ is examined, in particular, with respect to specification of H_0.

2. Methods

Spectral estimates of geophysical time series are often evaluated relative to a null spectral model based on an AR(1) process (e.g., Mann and Lees 1996; von Storch 1999). A number of more-detailed processes have also been proposed for representing null spectral models, including integrated and moving averages (Box et al. 2015; Klaus et al. 2015), use of a portion of the sample autocorrelation function (Priestley 1981; Garrido and 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 $\delta^{18}O$ during the last glacial has been unclear.

a. A null spectral model

A flexible model of H_0 is used here that admits for fitting the observed spectral estimate to variable detail. It is possible to represent any stationary stochastic process that has a continuous spectral density either as an autoregressive or moving-average (MA) process with a sufficiently high order (Priestley 1981). Whether GISP2 $\delta^{18}O$ is stationary is highly questionable, but, in practice, it is found that a combined AR and MA process allows for a parsimonious representation of H_0 for the GISP2 $\delta^{18}O$ spectral estimate. The ARMA model is

$$x(t) = \epsilon(t) + \sum_{i=1}^{p} a(i)x(t-i) + \sum_{i=1}^{q} m(i)\epsilon(t-i), \quad (1)$$

where $x(t)$ is the time series of interest, $a(i)$ are AR coefficients up to order p, and $m(i)$ are MA coefficients up to order q (Box et al. 2015). The $\epsilon(t)$ are the innovations 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, $H_0(p, q)$.

Equation (1) is fit in the time domain using an iterative maximum likelihood approach (Box et al. 2015). It is simplest to fit the data in the time domain in which they are originally provided, although methods exist for fitting ARMA models in either the time or frequency domain (Anderson 1977). For purposes of illustrating H_0, it is noted that the spectral representation of an ARMA process is

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

where δt is the time interval between observations in x, and s is frequency. ARMA processes can produce quasi-periodic

variability with peaked spectral representations, but such specifications should generally be avoided, in favor of smooth fits to the spectral estimates.

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 $H_0(p, q)$, an operation also referred to as prewhitening (Priestley 1981). 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. 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,

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

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

Equation (3) can be transformed using the natural logarithm of P,

$$f(\ln P) = \frac{P^k}{\Gamma(k)\theta^k} e^{-P/\theta}. \quad (4)$$

An advantage of using the natural logarithm is that the expected variance of $\ln P$ only depends on the degrees of freedom in the estimate, $\psi_3(k)$, where ψ_3 is the trigamma function. The mean of $\ln P$ is biased negative by $b = \psi_2(k) + \ln 2 - \ln 2k$, where ψ_2 is the digamma function, which becomes relevant when computing critical values.

A feature worth highlighting is that $\ln P$ is more normally distributed than P. Whereas the skewness of P is $(8/k)^{1/2}$, it is $\psi_4(k)/[\psi_3(k)]^{3/2}$ for $\ln P$ (Olshen 1938). The term 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 asymptotically approaching 0.5 as k becomes large. The variance of $\ln P$, v, is also less sensitive to concentrations of spectral energy than P. In section 4 it is shown that v is not especially biased by the presence of a moderate quasi-periodic contribution.

To gauge the adequacy of H_0, the ratio between the sample and expected variance is computed, $F = v/\psi_3$. Sample variance is $v = [1/(n-1)]\sum_{i=1}^{n}(\ln P_i - \overline{\ln P})^2$, where the sum is across n discrete frequencies, and the overbar indicates the sample mean. How close F is to unity provides a rule of thumb by which to select the order of the ARMA process. Specific tests for the goodness-of-fit of $\ln P$ to a log-gamma distribution could be applied such as the Kolmogorov–Smirnov or Cramer–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
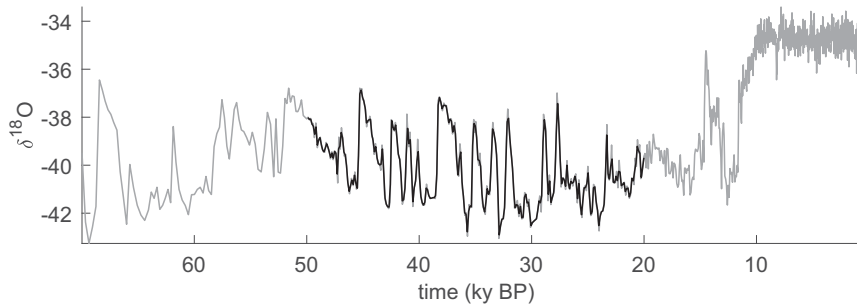
FIG. 1. Greenland Ice Sheet Project 2 δ^{18}O observations. Shown are both the original record at native sampling resolution (gray) and an interval from 50 to 20 ka that is interpolated to uniform resolution (black).

and demand a Monte Carlo evaluation of significance levels. Section 4 presents Monte Carlo realizations that indicate the plausibility of different values of F.

b. 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 multiple hypothesis tests. Thomson (1990) suggested using a significance level of 1/n when evaluating the full spectral estimate for periodic components. 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 for spectral peaks in a multitaper spectral estimate is approximately equal to the ratio of the frequency range to the bandwidth resolution. The frequency ranges from zero to the Nyquist frequency 1/(2Δt), and the bandwidth resolution of a multitaper spectral estimate is approximately $k/(2n\Delta t)$, where k is the number of tapers and n is the number of discrete positive frequencies. The number of independent tests is, thus, n/k, and the Bonferroni-adjusted significance level is

$$\alpha_b = \alpha k/n. \tag{5}$$

Following previous analyses (Schulz 2002; Schulz and 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. All tests are conducted using the Bonferroni correction unless specifically indicated otherwise.

Equation (5) requires no assumptions about dependencies between tests. An assumption that tests are not negatively correlated would also be appropriate and allows for a slightly more powerful test (Sidak 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 shows the δ^{18}O record from the Greenland Ice Sheet Project 2 (GISP2) with ages from stratigraphic layer counting (Meese et al. 1997). The focus is on GISP2 δ^{18}O within the 20- and 50-ka (ka indicates thousands of years ago) interval as representing a relatively homogeneous interval of the last glacial. The δ^{18}O samples have an average spacing of $\Delta t = 124$ yr between 20 and 50 ka, with the time between neighboring pairs of data points ranging between 77 and 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 years by convolving with an 11-point Hamming window, and, finally, decimated to 124-yr resolution to maintain the same average sampling interval. Interpolation to high-resolution followed by smoothing 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 and Mudelsee 2002).

Applying multitaper spectral estimation with three tapers, $k = 3$, gives a spectral estimate having the previously noted (Yiou et al. 1997; Wunsch 2000; Schulz and Mudelsee 2002; Ditlevsen et al. 2005) spectral peak centered at 1/(1470 yr) (Fig. 2). Dependencies in time series can reduce the effective degrees of freedom in a spectral estimate, but 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 kyr in the spectral estimate associated with the GISP2 δ^{18}O record averages 5.3 according to the estimation algorithm given by Percival and Walden (1993), and the effective degrees of freedom is 5.95 after leveling according 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 equaling $2k$ in leveled time series.
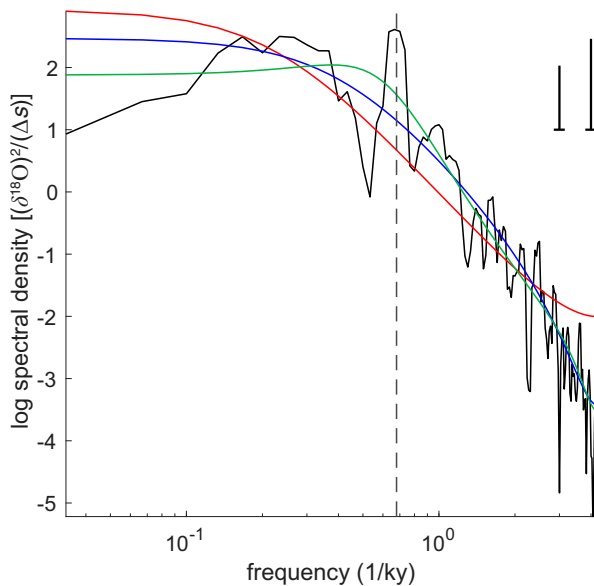
FIG. 2. Multitaper spectral estimate of the GISP2 $\delta^{18}O$ record between 50 and 20 ka. The spectral estimate is made using three tapers (black line; $k = 3$), and units are the natural logarithm of spectral power density. Candidate null spectral models H_0 are fit using autoregressive processes of order p and moving-average processes of order q, or $H_0(p, q)$. 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 $\alpha = 0.01$ critical value is shown assuming a single test (left vertical black line, shorter) and a multiple hypothesis test [right vertical black line, using Eq. (5)].

All combinations of H_0 involving p from 1 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 the 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 (Fig. 2). Note that ARMA processes can be specified that lead to quasi-periodic variability and peaked spectral estimates but that these should generally be avoided when specifying H_0.

The critical value of a properly leveled spectral estimate is constant. For a single hypothesis test with $k = 3$ and $\alpha = 0.01$ the critical value is 1.03 above the mean, and 1.46 above the mean for a multiple hypothesis test with $\alpha = 0.01$ and $n = 127$ frequencies [Eq. (5)]. The critical values for $\alpha = 0.05$ and single and multiple hypothesis tests are, respectively, 0.74 and 1.30. Units are suppressed by removing the sample mean, $\ln P' = \ln P - \overline{\ln P} + b$, as can be seen more clearly in the equivalent statement that includes only ratios of P,

$\ln P' = \ln(P/\overline{P}) - \overline{\ln(P/\overline{P})} + b$. The constant b corrects for the bias associated with taking the mean of the logarithm of P (see section 2). Insomuch as H_0 is adequate, $\ln P'$ is expected to follow a log-gamma distribution centered on zero. 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 critical values computed under the log-gamma assumption and is reflected in the fact that the variance of the leveled spectral estimate, $\nu = 0.87$, leads to $F = 2.20$ (Fig. 3b).

Leveling according to $H_0(1, 1)$ gives a spectral estimate whose variance is close to that expected, $F = 0.97$, and the spectral peak at 1/(1470 yr) is then insignificant at the $\alpha = 0.01$ level. Under $H_0(3, 2)$, F is 0.88 and the 1/(1470 yr) is insignificant even at $\alpha = 0.05$. More generally, values of F are between 0.86 and 1.06 for all examined version of H_0 with the notable exception of $H_0(1, 0)$, which has $F = 2.2$, and the minor exception of $H_0(5, 5)$, which has $F = 0.84$. 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. Simulations

Simulations are useful for purposes of quantifying the implications of 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 $\delta^{18}O$ record. On average, 6.6% of the spectral energy in $H_0(1, 1)$ resides within $1/1470 \pm 0.1/1470 \text{ yr}^{-1}$. H_1 is formed by adding a sinusoid at the 1/(1470 yr) frequency with an amplitude such that, on average, 22.4% of the spectral energy resides within $1/1470 \pm 0.1/1470 \text{ yr}^{-1}$, similar to that found in the observations (Fig. 2). In all, 10^4 realizations of H_0 and H_1 are each evaluated using versions of H_0 that are underspecified, perfect, and, overspecified. For each trial, the spectral estimate at the 1/(1470 yr) frequency and the maximum across frequencies are recorded, where the latter is the focus on account of being in a multiple hypothesis testing regime.

Leveling the simulated data using an underspecified model, $H_0(1, 0)$, leads to a 61% false rejection rate when assessing the significance of the maximum spectral estimate at a supposed $\alpha = 0.01$ level. The false rejection rate is illustrated in Fig. 4a as the mass of the H_0 distribution (black curve) that is to the right of the theoretical 1% critical value (1.46 logs spectral density; orange vertical line). Given that the target false rejection rate is 1%, $H_0(1, 0)$ is deeply flawed. The origin of this bias is traced to the fact that the spectral energy density in the vicinity of 1/(1470 yr) tends to be higher in the ARMA(1, 1) fit than accounted for by the ARMA(1, 0) fit, leading to $H_0(1, 0)$ being shifted toward positive values (dashed black line in Fig. 4a). $H_0(1, 0)$ also results in a large
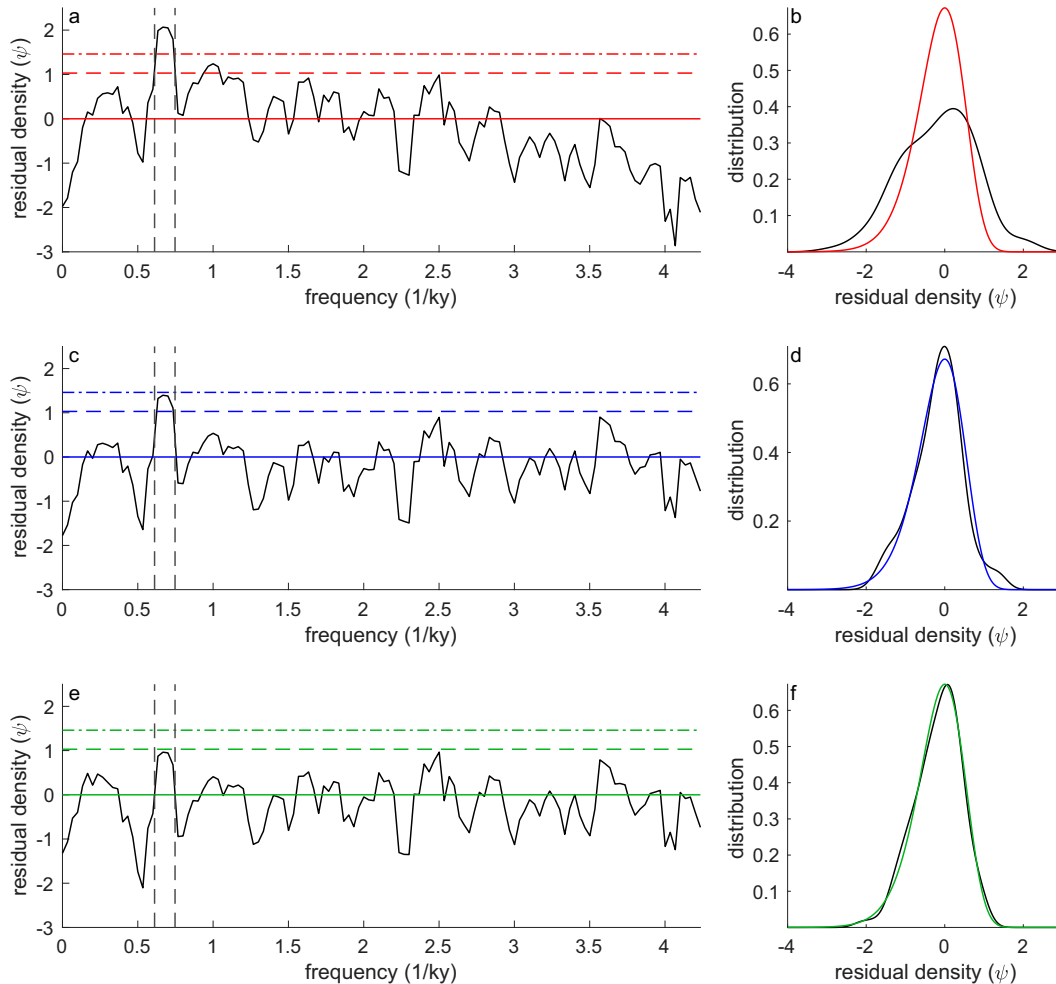
FIG. 3. Leveled spectral estimates indicating different levels of significance. The GISP2 δ^{18}O record shown in Fig. 1 is leveled using the three versions of H_0 depicted in Fig. 2. Leveled spectral density is expected to follow a log-gamma distribution [Eq. (4)] with variance 0.39 [$\psi_3(k = 3)$]. (a) Leveling according to $H_0(1, 0)$ indicates that the 1/(1470 yr) spectral peak is significant at the $\alpha = 0.01$ (dash–dotted line) significance level, but (b) the sample variance is $\upsilon = 0.87$. (c),(d) Here, $H_0(1, 1)$ gives $\upsilon = 0.38$ and indicates the 1/(1470 yr) peak is insignificant at $\alpha = 0.01$ (dash–dotted line) and significant at $\alpha = 0.05$ (dashed line). (e),(f) Here, $H_0(3, 2)$ gives $\upsilon = 0.35$ and the 1/(1470 yr) peak is insignificant.

spectral variance after leveling and F is distributed with an average of 2.15 (red vertical line in Fig. 4b) and 95th-percentile range of 1.40–3.25. This distribution of F is consistent with the value of $F = 2.20$ obtained when using $H_0(1, 0)$ with the actual observations (Fig. 3b), and supports the use of F as an effective metric for identifying inadequate versions of H_0. As opposed to assigning critical values assuming a log-gamma distribution (orange vertical line in Fig. 4a, $\alpha = 0.01$), it is also possible to assign critical values using the results of the simulations. Such an exercise indicates that the observed 1/(1470 yr) peak in the observations is insignificant at $\alpha = 0.01$ (i.e., the red vertical line is left of the black vertical line in Fig. 4a).

Use of a perfect model, $H_0(1, 1)$, results in a false rejection rate of 1.5%, similar to the intended rate of 1.0%. This similarity is reflected in the theoretical and empirical critical values being close, respectively equaling 1.46 (orange vertical

line in Fig. 4c) and 1.50 (black vertical line). These results also support the accuracy of the multiple hypothesis testing correction. The variance ratio is also very similar to the expected value of 1, with F averaging 1.03 (Fig. 4d). Furthermore, the mean of F under $H_1(1, 1)$ is 1.21, or only somewhat larger than under H_0, indicating that the variance metric is not overly sensitive to the presence of a moderate periodic component. If P was evaluated, as opposed to $\ln P$, the F associated with $H_1(1, 1)$ would, on average, be approximately double that under $H_0(1, 1)$, thereby having the undesirable feature of making evaluation of H_0 more sensitive to whether a periodic component is present.

Use of $H_0(3, 2)$ leads to an overspecified model of the time series. One result of overspecification is that the false rejection rate is low at 0.4% (Fig. 4e), with the empirical critical value (black vertical line) somewhat smaller than the
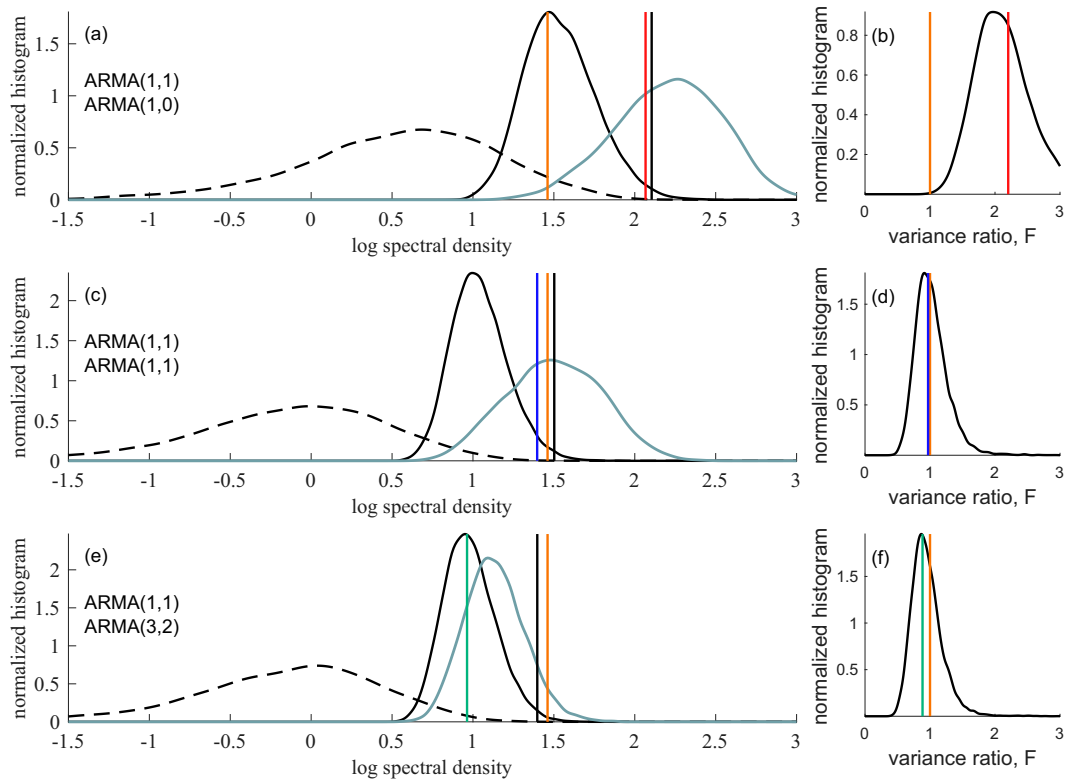
FIG. 4. Spectral analysis of synthetic data using different ARMA leveling processes. Synthetic data are realized 10^4 times according to an ARMA(1, 1) process fit to the GISP2 record and analyzed using the multitaper method with three tapers. Time series are variously leveled using (a),(b) $H_0(1, 0)$; (c),(d) $H_0(1, 1)$; or (e),(f) $H_0(3, 2)$ processes. Two null distributions and one alternate are shown in (a), (c), and (e): the distribution of log spectral density at the $1/(1470$ yr) frequency (H_0-single; dashed black line), the maximum value of each spectral analysis (H_0; black line), and the maximum value of each spectral analysis after adding a periodic component at $1/(1470$ yr) (H_1; gray line). The distribution of F, defined as the variance of the logarithm of leveled spectral estimates divided by the expected variance, is shown in (b), (d), and (f). The 1% critical value for H_0 is shown as inferred from synthetic realizations [vertical black lines in (a), (c), and (e)] and computed assuming that leveled spectra are distributed according to Eqs. (4) and (5) with $k = 3$ [vertical orange lines in (a), (c), and (e)]. The expected value of $F = 1$ is also shown [vertical orange lines in (b), (d), and (f)]. For purposes of comparison, log spectral energy density and values of F obtained from analyses of GISP2 $\delta^{18}O$ observations are indicated using colors that correspond to those in Figs. 2 and 3: $H_0(1, 0)$ (red), $H_0(1, 1)$ (blue), and $H_0(3, 2)$ (green).

theoretical critical value (orange vertical line). Another consequence of overspecification is removal of energy from the spectral peak in H_1, such that $H_0(3, 2)$ (black curve in Fig. 4e) and $H_1(3, 2)$ (gray curve in Fig. 4e) substantially overlap, yielding a low statistical power. These results do not, however, necessarily imply that $H_0(1, 1)$ is a better choice than $H_0(3, 2)$ for the actual GISP2 $\delta^{18}O$ data. If the simulated data are instead made from an ARMA(3, 2) process fit to the GISP2 $\delta^{18}O$ record, $H_0(1, 1)$ is strongly biased toward rejecting the null, with a false rejection rate of 22%, and $H_0(3, 2)$ then gives the least-biased results.

5. Further discussion

Some further remarks upon the selection of an appropriate ARMA model are appropriate. Criteria have been suggested elsewhere for selecting the order of ARMA processes, but these are generally concerned with minimizing residual variance after applying penalties for the number of predictor coefficients, for example, using Akaike's information criterion (Shibata 1976). Box and Pierce (1970) also gives 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. In the present case, the focus is on choosing an ARMA process that gives the correct residual variance such that F is close to unity.

The F metric is useful for ruling out ARMA processes that are underspecified but less so for processes that are overspecified. In the synthetic realizations presented in section 4, the 99th percentile of F under $H_0(1, 1)$ is 1.87, a value that is exceeded by 71% of the F realizations under $H_0(1, 0)$, indicating that underfitting using $H_0(1, 0)$ is typically identifiable (Figs. 4b,d). There is little distinction, however, between the

values of F obtained using $H_0(1, 1)$ and $H_0(3, 2)$ (Figs. 4d,f). The 1st percentile of F under $H_0(1, 1)$ is 0.58, a value that is greater than only 2.3% of the F realization associated with $H_0(3, 2)$. The relative insensitivity of F to using $H_0(1, 1)$ or $H_0(3, 2)$ reflects that the spectral variance explained by higher-order ARMA terms diminish rapidly in the current application.

Among the processes that give an F close to unity, there exist arguments for selecting either lower- or higher-order ARMA processes. Use of a low-order ARMA process tends to preserve statistical power to reject H_0, as is evident in the synthetic results (Fig. 4). The power of the test declines from 0.63 for $H_0(1, 0)$ to 0.50 for $H_0(1, 1)$ to only 0.08 for $H_0(3, 2)$, where power is measured as the mass of H_1 (gray curves in Figs. 4a,c,e) above the $\alpha = 0.01$ critical value (black vertical lines). Even using the synthetic $H_0(1, 0)$ results for which the greatest statistical power is found, however, leads to the conclusion that the observed 1/(1470 yr) spectral density in GISP2 $\delta^{18}O$ (red vertical line in Fig. 4a) is insignificant relative to the $\alpha = 0.01$ critical value (black vertical line in Fig. 4a; distinct from the theoretical but nonapplicable critical value in orange). Furthermore, results based on a low-order ARMA process may not adequately represent relevant processes. Indeed, any ARMA representation of GISP2 $\delta^{18}O$ is almost certainly underspecified in certain respects given the complex dynamics associated with North Atlantic climate variability and its recording in Greenland $\delta^{18}O$ (e.g., Guillevic et al. 2013; Rhines and Huybers 2014; Zuhr et al. 2021). In the present case it appears difficult to rule out the results of the higher-order versions of H_0.

Another consideration is that the presence of a true quasi-periodic contributions could lead to an overspecified formulation of H_0 through biasing v and, thereby, F high. Following an earlier suggestion for purposes of testing for structure in spectral estimates (Hannan 1961), an iterative approach can be taken whereby the most significant spectral peak is removed, F recomputed, H_0 refit if need be, and the process repeated until no further significant peaks are identified. In the present case, excluding log power density from a band within 1/(1470 yr) $\pm$ 0.2/(1470 yr) does not alter the conclusion that $H_0(1, 0)$ is inadequate, only reducing F from 2.20 to 1.91. Furthermore, higher-order ARMA processes remain near unity with $F = 0.84$ for $H_0(1, 1)$ and $F = 0.82$ for $H_0(2, 3)$ after removing the 1/(1470 yr) band.

A final consideration 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 $\delta^{18}O$ 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 and 50 ka, whereas the Holocene does not show variability at 1/(1470 yr). Other indicators of North Atlantic variability with approximately a 1500-yr 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.

6. Conclusions

A multitaper spectral estimate of GISP2 $\delta^{18}O$ 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, $H_0(1, 0)$ (Figs. 2 and 3a). $H_0(1, 0)$ is inadequate, however, as indicated by F being larger than expected (Fig. 3b) and simulations showing that the rate of false rejection is greater than 1 in 2 (Figs. 4a,b). Conversely, $H_0(1, 1)$ and $H_0(3, 2)$ indicate that F is consistent with expectations (Figs. 3d,f) and that the spectral peak is insignificant (Figs. 2 and 3c,e), with $p > 0.01$ for $H_0(1, 1)$ and $p > 0.05$ for $H_0(3, 2)$.

These spectral results are consistent with earlier findings based on time-domain analyses of the DO events that indicated a stochastic model as probably being the better description of DO event timing (Ditlevsen et al. 2005, 2007). The seeming prominence of the 1/(1470 yr) peak can be understood as a combination of a detailed spectral structure and the peak being selected for appearing unusually large. As opposed to being a clue that an underlying mechanism or external forcing imparts quasi-periodicity, it appears best to regard GISP2 $\delta^{18}O$ as entailing a spectral continuum of variability whose description requires an ARMA process higher than ARMA(1, 0) and probably higher than ARMA(1, 1). The ARMA techniques for evaluating the significance of spectral peaks presented here should also be applicable to evaluation of other geophysical records.

REFERENCES

Alley, R., S. Anandakrishnan, and P. Jung, 2001: Stochastic resonance in the North Atlantic. *Paleoceanography*, **16**, 190–198, https://doi.org/10.1029/2000PA000518.

Anderson, T., 1977: Estimation for autoregressive moving average models in the time and frequency domains. *Ann. Stat.*, **5**, 842–865, https://doi.org/10.1214/aos/1176343942.

Bond, G., and Coauthors, 1997: A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates. *Science*, **278**, 1257–1266, https://doi.org/10.1126/science.278.5341.

Box, G. E., and D. A. Pierce, 1970: Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Stat. Assoc.*, **65**, 1509–1526, https://doi.org/10.1080/01621459.1970.10481180.

——, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, 2015: *Time Series Analysis: Forecasting and Control.* John Wiley and Sons, 712 pp.

Braun, H., M. Christl, S. Rahmstorf, A. Ganopolski, A. Mangini, C. Kubatzki, K. Roth, and B. Kromer, 2005: Possible solar origin of the 1,470-year glacial climate cycle demonstrated in a coupled model. *Nature*, **438**, 208–211, https://doi.org/10.1038/nature04121.

Corrick, E. C., and Coauthors, 2020: Synchronous timing of abrupt climate changes during the last glacial period. *Science*, **369**, 963–969, https://doi.org/10.1126/science.aay5538.

Dansgaard, W., H. Clausen, N. Gundestrup, C. Hammer, S. Johnsen, P. Kristinsdottir, and N. Reeh, 1982: A new Greenland deep ice core. *Science*, **218**, 1273–1277, https://doi.org/10.1126/science.218.4579.1273.

Ditlevsen, P. D., M. S. Kristensen, and K. K. Andersen, 2005: The recurrence time of Dansgaard–Oeschger events and limits on the possible periodic component. *J. Climate*, **18**, 2594–2603, https://doi.org/10.1175/JCLI3437.1.

——, K. K. Andersen, and A. Svensson, 2007: The DO-climate events are probably noise induced: Statistical investigation of the claimed 1470 years cycle. *Climate Past*, **3**, 129–134, https://doi.org/10.5194/cp-3-129-2007.

Erhardt, T., E. Capron, S. O. Rasmussen, S. Schüpbach, M. Bigler, F. Adolphi, and H. Fischer, 2019: Decadal-scale progression of the onset of Dansgaard–Oeschger warming events. *Climate Past*, **15**, 811–825, https://doi.org/10.5194/cp-15-811-2019.

Garrido, J., and J. García, 1992: Periodic signals in Spanish monthly precipitation data. *Theor. Appl. Climatol.*, **45**, 97–106, https://doi.org/10.1007/BF00866398.

Goff, J. A., 2020: "Empirical prewhitening" spectral analysis detects periodic but inconsistent signals in abyssal hill morphology at the southern East Pacific rise. *Geochem. Geophys. Geosyst.*, **21**, e2020GC009261, https://doi.org/10.1029/2020GC009261.

Guillevic, M., and Coauthors, 2013: Spatial gradients of temperature, accumulation and δ^{18}O-ice in Greenland over a series of Dansgaard–Oeschger events. *Climate Past*, **9**, 1029–1051, https://doi.org/10.5194/cp-9-1029-2013.

Hannan, E., 1961: Testing for a jump in the spectral function. *J. Roy. Stat. Soc.*, **23B**, 394–404, https://doi.org/10.1111/j.2517-6161.1961.tb00421.x.

Johnsen, S., and Coauthors, 1992: Irregular glacial interstadials recorded in a new Greenland ice core. *Nature*, **359**, 311–313, https://doi.org/10.1038/359311a0.

Klaus, J., K. P. Chun, and C. Stumpp, 2015: Temporal trends in δ^{18}O composition of precipitation in Germany: Insights from time series modelling and trend analysis. *Hydrol. Processes*, **29**, 2668–2680, https://doi.org/10.1002/hyp.10395.

Li, C., D. S. Battisti, D. P. Schrag, and E. Tziperman, 2005: Abrupt climate shifts in Greenland due to displacements of the sea ice edge. *Geophys. Res. Lett.*, **32**, L19702, https://doi.org/10.1029/2005GL023492.

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

Mayewski, P. A., L. D. Meeker, M. S. Twickler, S. Whitlow, Q. Yang, W. B. Lyons, and M. Prentice, 1997: Major features and forcing of high-latitude Northern Hemisphere atmospheric circulation using a 110 000-year-long glaciochemical series. *J. Geophys. Res.*, **102**, 26 345–26 366, https://doi.org/10.1029/96JC03365.

Meese, D., and Coauthors, 1997: The Greenland Ice Sheet Project 2 depth-age scale: Methods and results. *J. Geophys. Res.*, **102**, 26 411–26 423, https://doi.org/10.1029/97JC00269.

Obrochta, S. P., H. Miyahara, Y. Yokoyama, and T. J. Crowley, 2012: A re-examination of evidence for the North Atlantic "1500-year cycle" at Site 609. *Quat. Sci. Rev.*, **55**, 23–33, https://doi.org/10.1016/j.quascirev.2012.08.008.

Olshen, A. C., 1938: Transformations of the Pearson type III distribution. *Ann. Math. Stat.*, **9**, 176–200, https://doi.org/10.1214/aoms/1177732309.

Percival, D. B., and A. T. Walden, 1993: *Spectral Analysis for Physical Applications.* Cambridge University Press, 583 pp.

Priestley, M. B., 1981: *Spectral Analysis and Time Series: Probability and Mathematical Statistics.* Academic Press, 943 pp.

Rahmstorf, S., 2003: Timing of abrupt climate change: A precise clock. *Geophys. Res. Lett.*, **30**, 1510, https://doi.org/10.1029/2003GL017115.

Rhines, A., and P. J. Huybers, 2014: Sea ice and dynamical controls on preindustrial and last glacial maximum accumulation in central Greenland. *J. Climate*, **27**, 8902–8917, https://doi.org/10.1175/JCLI-D-14-00075.1.

Rosen, J. L., E. J. Brook, J. P. Severinghaus, T. Blunier, L. E. Mitchell, J. E. Lee, J. S. Edwards, and V. Gkinis, 2014: An ice core record of near-synchronous global climate changes at the Bølling transition. *Nat. Geosci.*, **7**, 459–463, https://doi.org/10.1038/ngeo2147.

Sadatzki, H., and Coauthors, 2019: Sea ice variability in the southern Norwegian Sea during glacial Dansgaard-Oeschger climate cycles. *Sci. Adv.*, **5**, eaau6174, https://doi.org/10.1126/sciadv.aau6174.

Schulz, M., 2002: On the 1470-year pacing of Dansgaard-Oeschger warm events. *Paleoceanography*, **17**, 1014, https://doi.org/10.1029/2000PA000571.

——, and M. Mudelsee, 2002: REDFIT: Estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. *Comput. Geosci.*, **28**, 421–426, https://doi.org/10.1016/S0098-3004(01)00044-9.

Shibata, R., 1976: Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, 117–126, https://doi.org/10.1093/biomet/63.1.117.

Sidak, Z., 1967: Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Stat. Assoc.*, **62**, 626–633, https://doi.org/10.2307/2283989.

Stocker, T. F., and L. A. Mysak, 1992: Climatic fluctuations on the century time scale: A review of high-resolution proxy data and possible mechanisms. *Climatic Change*, **20**, 227–250, https://doi.org/10.1007/BF00139840.

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

——, 1990: Quadratic-inverse spectrum estimates: Applications to palaeoclimatology. *Philos. Trans. Roy. Soc.*, **A332**, 539–597, https://doi.org/10.1098/rsta.1990.0130.

Timmermann, A., H. Gildor, M. Schulz, and E. Tziperman, 2003: Coherent resonant millennial-scale climate oscillations triggered by massive meltwater pulses. *J. Climate*, **16**, 2569–2585, https://doi.org/10.1175/1520-0442(2003)016<2569:CRMCOT>2.0.CO;2.

Vaughan, S., 2005: A simple test for periodic signals in red noise. *Astron. Astrophys.*, **431**, 391–403, https://doi.org/10.1051/0004-6361:20041453.

——, R. Bailey, and D. Smith, 2011: Detecting cycles in stratigraphic data: Spectral analysis in the presence of red noise. *Paleoceanography*, **26**, PA4211, https://doi.org/10.1029/2011PA002195.

von Storch, H., 1999: Misuses of statistical analysis in climate research. *Analysis of Climate Variability*, H. von Storch and A. Navarra, Eds., Springer, 11–26.

Wolff, E. W., J. Chappellaz, T. Blunier, S. O. Rasmussen, and A. Svensson, 2010: Millennial-scale variability during the last glacial: The ice core record. *Quat. Sci. Rev.*, **29**, 2828–2838, https://doi.org/10.1016/j.quascirev.2009.10.013.

Wunsch, C., 2000: On sharp spectral lines in the climate record and the millennial peak. *Paleoceanography*, **15**, 417–424, https://doi.org/10.1029/1999PA000468.

Yiou, R., K. Fuhrer, L. Meeker, J. Jouzel, S. Johnsen, and P. A. Mayewski, 1997: Paleoclimatic variability inferred from the spectral analysis of Greenland and Antarctic ice-core data. *J. Geophys. Res.*, **102**, 26 441–26 454, https://doi.org/10.1029/97JC00158.

Zuhr, A. M., T. Münch, H. C. Steen-Larsen, M. Hörhold, and T. Laepple, 2021: Local-scale deposition of surface snow on the Greenland ice sheet. *Cryosphere*, **15**, 4873–4900, https://doi.org/10.5194/tc-15-4873-2021.