RESEARCH ARTICLE

Frequency selection in paleoclimate time series: A model-based approach incorporating possible time uncertainty

Peter M. Franke[1] | Brian Huntley[2] | Andrew C. Parnell[3]

[1]School of Mathematics and Statistics, University College Dublin, County Dublin, Ireland

[2]School of Biological and Biomedical Sciences, Durham University, County Durham, U.K.

[3]School of Mathematics and Statistics, and Insight Centre for Data Analytics, University College Dublin, County Dublin, Ireland

Correspondence
Peter M. Franke, School of Mathematics and Statistics, University College Dublin, Dublin 4, Belfield, Ireland.
Email: peter.franke@ucdconnect.ie

A key aspect of paleoclimate time series analysis is the identification of frequency behavior. Commonly, this is achieved by calculating a power spectrum and comparing this spectrum with that of a simplified model. Traditional hypothesis testing method can then be used to find statistically significant peaks that correspond to different frequencies. Complications occur when the data are multivariate or suffer from time uncertainty. In particular, the presence of joint uncertainties surrounding observations and their timing makes traditional hypothesis testing impractical.

In this paper, we reexpress the frequency identification problem in the time domain as a variable selection model where each variable corresponds to a different frequency. We place this problem in a Bayesian framework that allows us to place shrinkage prior distributions on the weighting of each frequency, as well as include informative prior information through which we can take account of time uncertainty.

We validate our approach with simulated data and illustrate it with analysis of mid- to late Holocene water table records from two sites in Northern Ireland—Dead Island and Slieveanorra. Both case studies also show the extent of the challenges that researchers may face. We therefore present one case that shows a good model fit with a clear frequency pattern and the other case where the identification of frequency behavior is impossible. We contrast our results with that of the extant methodology, known as REDFIT.

KEYWORDS
frequency analysis, parameter selection, REDFIT, time uncertainty

1 | INTRODUCTION

Frequency analysis of cyclic and quasicyclic dynamics in paleoclimate time series plays a vital role in identifying the causes of paleoclimatic variations and thus in disentangling the mechanisms underlying the processes of these variations. Identifying frequencies characteristic of solar activity, and variations in the Earth's and lunar orbits, in paleoclimate time series has led to the recognition of the importance of at least orbital variability and solar variability as drivers of global climatic changes. The primary example is the demonstration of the orbital frequencies in Quaternary paleoclimatic time series that span multiple glacial–interglacial cycles (Hays, Imbrie, & Shackleton, 1976). Another example is the

identification of solar activity periodicities in the record of last glacial climatic variability (Bond et al. 1997). Warner and Domack (2002) identified frequencies in their paleoclimate time series from Antarctica that correspond to solar variability and lunar orbital periodicities.

In this paper, we introduce a new method to identify frequency behavior while dealing with time irregularity and uncertainty. The most commonly used methods in the assessment of frequency patterns are based on Fourier methods that rely on a discretely observed, evenly spaced time variable. In the case where data are continuous in time and are unevenly spaced, Lomb–Scargle methods are often used (Lomb, 1976; Scargle, 1982, 1989). However, in paleoclimate time series, data are not only irregularly spaced but also often observed with known time uncertainty.

A commonly used package for dealing with frequency analysis of irregular data is that of REDFIT (Mudelsee, 2009; Schulz & Mudelsee, 2002). This method involves calculating an empirical Lomb–Scargle periodogram for the time series in question and then comparing this empirical periodogram with quantiles of the theoretical periodogram of an AR(1) process (Schulz & Mudelsee, 2002). At frequency levels where the empirical periodogram is larger than that of the theoretical quantiles, the frequency is adjudged to have attained "statistical significance". A number of issues with this model persist:

1. To take account of time uncertainty, a number of bootstrap simulations are required to resample the time series. This is computationally expensive.
2. Because the empirical periodogram is only compared with that of an AR(1), other models that may be preferable are ignored.
3. Care needs to be taken when choosing the number of frequencies to test, because this will increase the probability of a type 1 error in the hypothesis test.
4. The statistical significance framework provides that all null hypotheses will be rejected if the data set is large enough. This will lead to the overestimation of the number of important frequencies.

In our approach, we reexpress the frequency analysis problem as that of a Bayesian variable selection in the time domain, where we explore a large class of model behavior exhibiting both frequency and nonfrequency patterns and allow the data to choose between them through the model fitting process. Our method allows for time uncertainty to be naturally incorporated as part of the Bayesian prior distribution, and we have explicit probabilities of frequencies being included in the model.

Temporal uncertainty arises because it is impossible to obtain the exact age of each layer in a fossil core. Ages of certain layers can be obtained through, for example, radiocarbon dating, but these are uncertain due to both laboratory measurement error and the uncertainty associated with calibrating the laboratory measurement into calendar years. This uncertainty is further compounded because other layers also need dating, and so, an interpolation strategy is required. We use Bchron (Haslett & Parnell, 2008; Parnell & Gehrels, 2015) that provides ages and standard errors for the timing of each data point.

Our paper is structured as follows. In Section 2, we discuss how frequency analysis has traditionally been performed with palaeoclimate time series. In Section 3, we outline our new approach and introduce all notations. In Section 4, we provide a simple simulated example and show that it outperforms REDFIT at identifying frequencies. In Section 5, we apply our new models to water table time series from two sites in Northern Ireland—Dead Island and Slieveanorra. We conclude and suggest avenues for further research in Section 6.

2 | TRADITIONAL APPROACH TO FREQUENCY IDENTIFICATION

A discretely sampled time series can be decomposed into a constant, noncyclical trend function and a sum of cyclical components via the following form:

$$y(t_i) = g(t_i) + \sum_{j=1}^{k} \beta_j \cos(2\pi f_j t_i) + \sum_{j=1}^{k} \gamma_j \sin(2\pi f_j t_i) + \xi_i. \tag{1}$$

Here,

1. $y(t_i)$ is a univariate observation (e.g., water table records, transformed pollen counts, etc.) sampled at time t_i.
2. t_i represents the continuous time of the ith observation.

3. $g(t_i)$ is a noncyclical trend function of known form usually also containing a constant term.
4. k is the number of frequency components.
5. f_j is the frequency for component j.
6. β_j and γ_j represent the Fourier coefficients of cosine and sine functions at the jth frequency, respectively.
7. ξ_i is the residual error term, commonly assumed to be normally distributed $\xi_i \sim N(0, \sigma^2)$.

The power spectral density $P(f)$ is obtained based on the estimated coefficients of the cyclical components β and γ in Equation (1) as follows:

$$P(f_j) = \frac{\beta_j^2 + \gamma_j^2}{2}. \tag{2}$$

In general, obtaining power spectral densities can be achieved either directly from the data without applying any intermediate steps (e.g., fast Fourier transform) or indirectly, which requires estimating Fourier coefficients β and γ from a version of Equation (1) in the first step and obtaining power spectra afterwards (Equation (2)).

Below, we present a few of the most relevant methods of frequency analysis and discuss their applicability in the analysis of irregular and uncertain time series. We discuss the main features of the traditional methods to estimate frequency behavior in the context of the general decomposition function presented in Equation (1). We use the notation of Equation (1) for the majority of the paper.

2.1 | Spectra of the frequency domain method applicable to unevenly spaced time series

The Lomb–Scargle method (Lomb, 1976; Scargle, 1982, 1989) is a transformation of Equation (1) with exclusion of the same elements as in the case of the discrete Fourier transform. It is a direct method of frequency analysis that evaluates spectra at angular frequencies $\omega_k = 2\pi f_k > 0$. It also incorporates an adjustment term κ to account for the possible irregular timing of observations. The Lomb–Scargle power spectrum is defined as follows:

$$P(\omega_k) = \frac{1}{2\sigma_y^2} \left\{ \frac{\left[\sum_{i=1}^N (y_i - \bar{y}) \cos \omega_k(t_i - \kappa) \right]^2}{\sum_{i=1}^N \cos^2 \omega_k(t_i - \kappa)} + \frac{\left[\sum_{i=1}^N (y_i - \bar{y}) \sin \omega_k(t_i - \kappa) \right]^2}{\sum_{i=1}^N \sin^2 \omega_k(t_i - \kappa)} \right\}, \tag{3}$$

where σ_y^2 is the variance of y (not ξ), and $\frac{1}{2\sigma_y^2}$ plays the role of a constant normalizing the power spectrum. The κ is defined as follows:

$$\kappa = \frac{1}{2\omega_k} \tan^{-1} \left[\frac{\sum_{i=1}^N \sin 2\omega_k t_i}{\sum_{i=1}^N \cos 2\omega_k t_i} \right]. \tag{4}$$

The bandwidth of the Lomb–Scargle method is not as restricted as in the case of the discrete Fourier transform due to irregular sampling. Mudelsee (2009) highlighted the difficulty of selecting the highest frequency and suggested extending the frequency range to 110% of the Nyquist critical frequency.

2.2 | Spectra of the time domain method

The harmonic regression (Vaníček, 1969) method estimates Fourier coefficients in Equation (1) by minimizing the sum of squared error terms $\sum \xi_i^2$. It can incorporate any functional form of a trend and is not bounded by minimum or maximum frequency and by frequency resolution (Vaníček, 1969). However, the parameter k determining the number of cyclical components in the model needs to be kept low, as the model has a tendency to overfit in high-dimensional problems. The power spectrum is estimated as usual using Equation (2).

2.3 | Frequency identification methods

Scargle (1982) proved that the normalized Lomb–Scargle spectrum follows an exponential distribution with mean 1 and that it can be tested with its quantile function as follows:

$$F^{-1}(\alpha^*) = -\log(1 - (1 - \alpha^*)^{1/k}), \tag{5}$$

indicating a level above which peaks in the spectrum are significant at the α^* level.

The REDFIT methodology is based on the Lomb–Scargle algorithm in Equation (3) (Schulz & Mudelsee, 2002), where it applies a different approach to frequency identification. REDFIT assumes that the null hypothesis for a stochastic process exhibits a continuous decrease of spectral amplitude with increasing frequency and models that behavior with the continuous time AR(1) process (Mudelsee, 2002), as follows:

$$y(t_i) = \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right) y(t_{i-1}) + \xi_i, \tag{6}$$

where

- τ is the time persistence scaling factor, and
- ξ_i is the heteroscedastic random component distributed as $\xi_i \sim N\left(0, 1 - \exp\left(-2\frac{t_i - t_{i-1}}{\tau}\right)\right)$.

Under this assumption, the confidence intervals of the AR(1) model, obtained via Monte Carlo simulation and being equivalent to the χ^2 multiple frequency test, serve as an indicator of significant frequencies in the data spectrum. The REDFIT methodology (Schulz & Mudelsee, 2002) after Thompson (1990) suggests applying a false-alarm level G via the following:

$$G = \left(1 - \frac{1}{n^*}\right) \times 100\%, \tag{7}$$

where n^* is the number of data points in each (possibly overlapping) segment of the data (Mudelsee, 2009). Other approaches to setting the false alarm also exist (e.g., Siegel's test; Siegel, 1980). The currently available version of the RED-FIT package (version 38e) does not support custom false-alarm levels and instead provides results for 80%, 90%, 95%, and 99% significance levels.

Applying the AR(1) model for the purpose of identifying "significant" frequencies has a number of issues, as follows:

1. It limits the applicability of the REDFIT methodology to the cases where AR(1) noise is the only valid competitor to full frequency behavior, thus disregarding the possibility of intermediate complexity models.
2. The time persistence parameter τ of the AR(1) model, upon which the shape of the spectrum strongly depends, is treated as fixed. In fact, it is estimated from the data as $\hat{\tau}$ and is thus a random variable. Consequently, there may be considerable uncertainty over the level of the spectrum due to the epistemic uncertainty in τ.

The REDFIT package also allows us to apply one of the five different window functions: rectangular, Welch, Hanning, Triangular, and Blackman–Harris (Schulz & Mudelsee, 2010) to reduce the impact of spectral leakage. Although reduction of spectral leakage is a very desired effect, the choice of a window function influences the shape of spectral densities and amplitudes of spectral peaks (Trauth, 2015)[1]. Using time-uncertain water table records from Dead Island and Slieveanorra, we demonstrate in our case study and in Appendix A that the choice of different window functions influences the results of hypothesis testing in the REDFIT model.

2.4 | Time uncertainty

Estimating the timing of observations from paleoclimatic data is typically completed via dating methods (e.g., radiocarbon; Bowman 1990). However, incorporating the time uncertainty that arises from such methods is very challenging and so often ignored (e.g., Essefi, Mefteh, Medhioub, & Yaich, 2014; Swindles, Patterson, Roe, & Galloway, 2012). Traditional methods of frequency analysis do not account for the uncertainty in time, as the problem is not embedded within the Lomb–Scargle and harmonic regression models. In Section 4, we demonstrate that ignoring time uncertainty can have grave implications for frequency identification.

One method to incorporate time uncertainty (as used by Mudelsee et al., 2009) is to model time-scale errors parametrically to determine upper levels of the AR(1) spectrum and estimate uncertainty of spectral peaks. Another method is to jitter time estimates multiple times outside of a frequency model, evaluate spectral densities for each realization, and merge the results (Rhines & Huybers, 2011).

Our alternative approach is to include time uncertainty directly through the prior distribution, corresponding to a standard errors-in-variables structure (e.g., Dey, Ghosh, & Mallick, 2000). We present this in the next section.

---

[1]The reference describes the application of window functions in evenly sampled time series; we did not find any publications discussing the influence of window functions on frequency identification in the domain of time-uncertain time series.

3 | BAYESIAN FREQUENCY SELECTION

In this section, we outline our main contribution, a model-based Bayesian approach, to frequency identification. We term the model we use as *Bayesian frequency selection* (BFS); it can be considered part of the class of shrinkage and selection methods that aim to choose a subset of predictors (here, frequencies) as being important for explaining the variability in the response. Many other shrinkage and selection methods have been proposed (e.g., lasso, ridge regression, etc.; for a more complete list, see Hastie, Tibshirani,& Friedman, 2009), although application to the problem of frequency selection is limited. A critical assumption of such models, which we follow here, is that a large number of frequencies included in, for example, the harmonic regression model, are negligible so that their regression coefficients β_j and γ_j are all 0. In a Bayesian context, many of the aforementioned shrinkage models do not shrink parameter values fully to zero (Hastie et al. 2009). In our BFS model, we use a version of Bayesian shrinkage that does set parameters at 0, following the method of George and McCulloch (1993, 1997).

Our proposal is to apply the BFS to frequency identification by introducing indicator variables I_j into Equation (1). The indicator variables take the value 0 or 1 with a predefined prior Bernoulli distribution so that $I_j \sim Bernoulli(q)$, where q indicates the probability that frequency j is important in explaining the variability in the response (Kuo & Mallick, 1988). In particular, we force the indicator variables I_j to assume the same values for both sine and cosine terms at each particular frequency jointly. We thus adapt Equation (1) slightly to give the following:

$$y(t_i) = g(t_i) + \sum_{j=1}^{k} I_j \beta_j \cos(2\pi f_j t_i) + \sum_{j=1}^{k} I_j \gamma_j \sin(2\pi f_j t_i) + \xi_i. \tag{8}$$

A key aspect is that we can include external information about the parameters (e.g., from other data sources) by specifying their prior probability distributions. In the context of frequency analysis, we aim to obtain the posterior distribution of the parameters that govern the cyclical components, namely β_j and γ_j. The central contribution of our approach is the specification of a model through which the important frequencies are chosen automatically.

Through our BFS approach, we can still specify the total number of frequencies k to be large and let the inference procedure (constrained by a prior distribution) choose the number of important frequencies. The lack of such a restriction in k allows us to analyze a very wide spectra in search of important components. The harmonic regression model in many such cases would fail to recognize the true pattern and typically would identify too many frequencies.

Another benefit of our proposal is that within this framework, we can incorporate any functional form of trend $g(t_i)$ as appropriate to the problem at hand. In contrast, the frequency domain methods are not designed to cope with trends. Spectra of time series containing such components are distorted, and thus, traditionally, it has been necessary to remove them (Mudelsee, 2009). This can cause problems when low frequency components compete with the trend to explain variability.

In our application, we model the trend component $g(t_i)$ as a linear combination of polynomial basis functions. We set the maximum degree to be 3. This is of course an arbitrary choice, but values higher than this would induce their own cyclical behavior and so compete directly with the frequency component. The third degree polynomial term will have at most one complete cycle across the time support of the data, and so, the value of 3 seems appropriate. Furthermore, for computational purposes, we find it convenient to express Equation (8) in matrix form, as follows:

$$Y = X\alpha + Z_1 I \beta + Z_2 I \gamma + \xi, \tag{9}$$

where

- Y is the $n \times 1$ column vector of observed $y(t_i)$, $i = 1, \dots, n$ values;
- X is the $n \times (l + 1)$ standardized design matrix of the constant and trend components, where l is the number of trend components in a polynomial basis. Thus, the i, jth element of X is $t_i^{*,j-1} = \left(\frac{t_i - \bar{\bar{t}}}{\bar{s}_t}\right)^{j-1}$, where $\bar{\bar{t}}$ is the mean of means of timings of observations, and $\bar{s}_t$ is the mean of standard deviations of timings of observations—both are given as data. The standardization enables easier implementation of noninformative priors;
- α is the $(l + 1) \times 1$ column vector of the constant and trend parameters;
- Z_1 and Z_2 are the $n \times k$ matrices of cosine and sine terms, respectively. The i, jth element of Z_1 is $\cos(2\pi f_j t_i)$, and for Z_2, it is $\sin(2\pi f_j t_i)$;

- I is a $k \times k$ diagonal matrix of indicator variables with elements $I_j = 1$ or $I_j = 0$ on the diagonal and with zeros on all off-diagonals (note that we use J rather than I for the identity matrix);
- β and γ are $k \times 1$ vectors of parameters of the cosine (β) and sine (γ) terms; and
- ξ is $n \times 1$ is vector of the residual error terms with $\xi \sim N(0, \sigma^2 J)$. This is white noise Gaussian covariance matrix.

Taken together, we have parameters $\theta = \{\alpha, \beta, \gamma, \sigma^2, I, q\}$ and data D containing the design matrices $D = \{Y, X, Z_1, Z_2\}$. Note that the frequencies are included in the model via the matrices Z_1 and Z_2. Henceforth, in describing the model, we always use this form of the notation, discarding the earlier form from Equation (1) because it is more relevant to the explanation of the BFS model.

3.1 | Prior distributions

We now turn our attention to the choice of prior distributions for unknown parameters. Where possible, informative priors should be used to guide the values of these, as in most normal scientific experiments. Where not possible, we can use vague or weakly informative priors to enable valid model fits while avoiding undue influence on the identification of frequencies. The parameter that is most sensitive to prior choice is that of I_j and its Bernoulli associated probability q. We propose several choices:

1. In cases where there is no knowledge whatsoever about q, we might assume a vague prior, for example, that q is equally likely to take any value from 0% to 100%, so that the prior distribution is $q \sim U(0, 1)$.
2. In cases where the distribution of the parameter q requires more weight toward a point between 0% and 100%, one could assume the Beta prior distribution so that $q \sim Be(\alpha^*, \beta^*)$.
3. In cases where there is known to be a small number of important frequencies, we might redefine $q = \frac{m}{k}$, where m is the number of important frequencies, and k is the total number of frequencies (as in Equation (8)) set to $k = \frac{n \times ofac}{2}$, where the *ofac* parameter is the oversampling factor, by default *ofac* $= 1$. The value of m can then be fixed or given a prior. For example, if there are estimated to be a maximum of γ important frequencies, then, for example, $m \sim DU(0, \gamma)$, where DU is the discrete uniform distribution. By default, we set the parameter $\gamma = 10$.

For the remainder of the paper, we propose to use the latter formulation, as it most closely matches the real-world behavior we hope to model.

For the parameters associated with the trend and frequency components, we use vague priors, as follows:

$$\alpha \sim N\left(0, \sigma_\alpha^2 J\right), \ \beta \sim N\left(0, \sigma_\beta^2 J\right), \ \gamma \sim N\left(0, \sigma_\gamma^2 J\right), \tag{10}$$

where the prior variances are all fixed at large values. For the residual variance, we use the standard reference prior, as follows:

$$\sigma_\xi^2 \sim IG(e^*, f^*), \tag{11}$$

where e^* and f^* are set to small values.

To include time uncertainty, we simply add an extra layer to the model so that θ now includes t_i for $i = 1, \ldots, n$ and that D similarly includes $\mu_{t_i}, \sigma_{t_i}^2$. We write t as the full set of all unknown true times and μ_t, σ_t^2 as the set of means and variances, respectively. The prior distribution is as follows:

$$t_i \left| \mu_{t_i}, \sigma_{t_i}^2 \sim TN_{t_{j-1}}^{t_{j+1}}\left(\mu_{t_i}, \sigma_{t_i}^2\right), \right. \tag{12}$$

where

- t_i is the true and unknown age of observation i bounded at the extreme ends, t_0 and t_{n+1}, set to very small and very large values, respectively;
- μ_{t_i} is the mean age of observation i estimated from a chronology model; and
- $\sigma_{t_i}^2$ is the variance of the age of observation i, estimated at the same time as μ_{t_i}.

The extra computational issue is that now, the matrices Z_1 and Z_2 need to be changed for each new realization of t_j. However, this overhead is the necessary price to be paid for including time uncertainty.
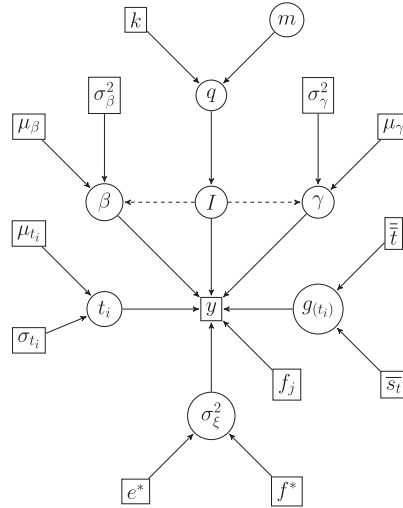
FIGURE 1 Directed acyclic graph of our Bayesian frequency selection model. Circles correspond to parameters to be estimated, whereas squares indicate data. Arrows indicate the direction of information flow

A more complete model would use posterior samples from Bchron, where each set of samples is not normally distributed but remains monotonically consistent. We found such an approach impossible to implement without computational challenges and leave this for future research.

In practice, these estimated ages are not normally distributed but Parnell and Gehrels (2015) have shown that, in a reasonable size of data sets, this difference is unimportant in the final model predictions. In our practical application, we use the chronology model Bchron (Haslett & Parnell, 2008) to obtain the posterior distribution of timing from each layer. In the next step, we approximate those distributions by Gaussian distributions by calculating the pairs $\mu_{t_i}, \sigma_{t_i}^2$. In the model, we impose truncation on the approximated realizations as per Equation (12).

3.2 | Model fitting

Our full posterior distribution is thus as follows:

$$p\left(\alpha, \beta, \gamma, \sigma^2, I, q, t, m | Y, X, Z_1, Z_2, \mu_t, \sigma_t^2\right), \tag{13}$$

which is separated into likelihood and prior, respectively, as follows:

$$p(Y|X, Z_1, Z_2, t, \sigma^2, \alpha, \beta, \gamma) \times p(\alpha)\, p(\beta)\, p(\gamma)\, p(\sigma^2)\, p(I|q)\, p(q|m)\, p(m)\, p\left(t|\mu_t, \sigma_t^2\right). \tag{14}$$

An alternative way to display the model fit is in the form of a directed acyclic graph (Figure 1) that shows the relations between parameters and data and also the conditional independencies.

3.2.1 | Sampling from the posterior distribution

We use the JAGS package (Plummer, 2017) for the purpose of obtaining posterior distributions of the parameters. It is now a standard method in the Bayesian modeling toolbox that works well even in high-dimensional settings such as ours. To provide insight into the implementation of the BFS model, we enclose our JAGS code in Appendix B.

The package uses a number of algorithms to generate samples from posterior distributions: Gibbs sampling, slice sampling, adaptive rejection sampling, and Metropolis–Hastings (Plummer, 2017). The sampling algorithm starts with the definition of the full posterior (Equation (13)) and, from some initial values of the parameters used to start the iterative process, proceeds by updating the parameters either individually or in block fashion. After a suitable period (known as the burn-in period), a set number of the parameter values are stored and thinned out to remove autocorrelation. For our algorithm, we keep 2,500 samples of the parameters from the posterior and remove 1,500 burn-in samples. We check the convergence using the standard Brooks–Gelman–Rubin diagnostics (Brooks & Gelman, 1998; Gelman & Rubin, 1992) to ensure that the parameter estimates are stable.

3.2.2 | Posterior post hoc calculations

Upon obtaining samples from the posterior distribution, we make inferences about the parameters of interest. In our case, the parameters of interest are that of the sine and cosine coefficients β_j and γ_j, respectively, and their relevant indicator variables I_j. From these values, we create two important quantities. The first is our weighted power spectrum created as a part of the BFS methodology, as follows:

$$P_{BFS}(f_j) = \frac{I_j \beta_j^2 + I_j \gamma_j^2}{2}, \tag{15}$$

which we can now estimate with uncertainty for all frequencies, as we have full access to the posterior distribution of the parameters. The second is the probability of a frequency at f_j, which we can calculate from the posterior mean of I_j. A high value (i.e., close to 1) indicates a likely important frequency. We recommend researchers to tailor their interpretation to the specific problem at hand and to use the tools of prior distributions to create a bespoke version of their analysis. In situations where we observe small posterior means of I_j and wide posterior intervals for β_j and γ_j, we would still argue that such frequencies are unimportant. Aside from our two main outputs, we also have access to all the posterior samples of the other parameters. We can thus determine the degree to which the trend components are important through α, assess the model fit via σ^2 (where small σ^2 indicates good model fit), or look at the posterior estimates of timing from t.

3.2.3 | Model cross-validation

We assess the predictive performance of the model by applying K-fold cross-validation. This allows us to determine the predictive fit of the model on data that has not been used as part of the fitting process and thus provides a much harsher test of model performance. We use standard 5-fold cross-validation even though our data are time series. Such an approach has been validated by Bergmeir and Benitez (2012).

In the presence of time uncertainty, the observations are of course not observed at known time points, so out-of-sample performance may be adversely affected, even when the model is identified correctly. In order to combat this problem and obtain a more accurate picture of cross-validation performance, we adjust the timing in Equation (12) of the data points by the output from our Bayesian model, namely the estimates of the t_i parameters. By doing so, we attempt to predict the true timing of actual observations instead of predicting the timing that a certain dating methodology provides. In our simulated data sets, this approach yields a fairer out-of-sample performance, giving a clearer picture of the model validation performance under time uncertainty.

4 | ACCURACY OF REDFIT AND BAYESIAN FREQUENCY SELECTION METHODS

In this section, we conduct a simulation experiment where we created 110 simulated data sets (100 with frequency behavior and 10 without) of varying complexities and fit both our BFS and the REDFIT model. Our aim is to see which of the models identifies the correct frequencies. Because both methods provide confidence levels about the frequencies (BFS via the posteriors of I_j, and REDFIT via the significance level), we use receiver operator characteristic (ROC) curves (Swets, 1996) to contrast their performance.

4.1 | Simulation and testing methodology

We test both methods using 110 simulated time-uncertain data sets with the following randomly generated:

- number of observations $N \sim DU(100, 150)$;
- 100 data sets with the number of frequencies in the signal $k \sim DU(1, 6)$ randomly assigned to either sine or cosine coefficients with amplitudes $\beta_j = \gamma_l = 10$, and 10 data sets, where $k = 0$;
- residual error term $\xi_i \sim N(0, \sigma^2)$, where $\sigma \sim U(2, 30)$;
- uniformly distributed true timing of observations $t_{\text{true},1:N} \sim U(0, 100)$, sorted in ascending order to model nonoverlapping attribute of paleoclimate time series;
- normally distributed timing of observations $t_{\text{observed}_i} \sim N\left(t_{\text{true}_i}, \sigma^2_{\text{observed}}\right)$, where $\sigma_{\text{observed}} \sim (0.2, 2.5)$ (note $\mu(dt_{\text{observed}}) \approx 0.8$); realizations are sorted in ascending order.

The objective of this specification is to mimic basic attributes of paleoclimate time series, which demonstrate substantial variety in frequency and noise behavior.

In conducting our tests of both frequency identification methods, we apply the same randomly generated 110 simulated time-uncertain time series to both methods. We use default settings in REDFIT and run 5,000 simulations using the BFS method with a burn-in of 20%. We identify the peaks above specified significance levels in the results obtained with REDFIT and probability levels in the BFS method as frequencies selected by those methods. Next, we compare those with the true frequencies in all data sets.

In identifying true positives (or false negatives), we accept the results if the selected frequencies are in (or outside) of a narrow range of ± 1 frequency f_j in frequency vectors generated with the highest resolution obtained without oversampling in the Nyquist range. We define this range to avoid misclassifications caused by frequency discretization and frequency leakage.

For the purpose of this test, we apply the default oversampling setting (ofac) of the REDFIT method $ofac_{REDFIT} = 4$, as it is commonly used by researchers. We apply no oversampling in the BFS method $ofac_{BFS} = 1$ because none is required, and by doing so we avoid slowing the model down considerably. To make a fair comparison between the two methods, and compare the results using the same range of frequency buckets, we define them as follows: let f_j be a frequency in discrete frequency vectors generated by either of the methods that is the closest to the true frequency f_k; then, the range of frequency buckets containing the true frequency is $r_j \subset [f_{j-ofac}, f_{j+ofac}]$. This should enable fair comparison despite the differences in oversampling between the methods.

In the next step, we plot the ROC curve (Figure 2) for the BFS method and add empirical true positive and false positive rates for the only four significance levels available in REDFIT (99%, 95%, 90%, and 80%). For REDFIT, this involves some degree of linear interpolation.

4.2 | Results from the simulation experiment

The area under the ROC curve generated by the BFS method is 86%, and given the extreme characteristics of the generated time-uncertain time series, we assess this result as very good. We cannot generate the entire area under the ROC curve using REDFIT; the available four points plotted in Figure 2 give a reference for comparison of the performances with the area under the ROC curve of 59%. It seems unlikely that, even if other significance levels were available to complete the full curve, the area under the ROC curve for REDFIT would approach that of the BFS method. In our view, the BFS method outperforms REDFIT in terms of frequency identification. It also allows us to apply a number of diagnostics supporting validation (such as cross-validation, convergence diagnostics, etc.).
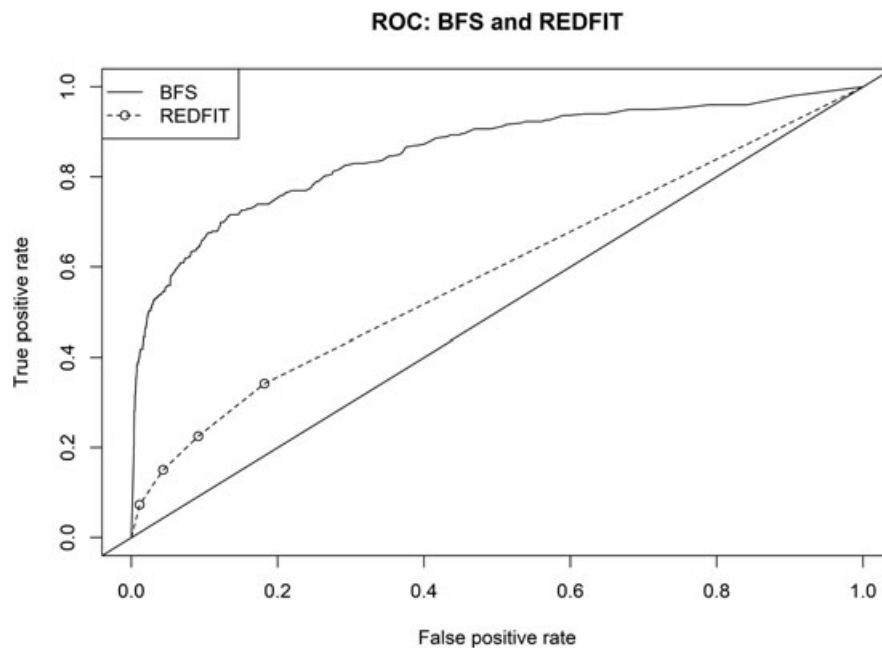
FIGURE 2 ROC curves for BFS and REDFIT methods based on 110 simulated data sets. ROC = receiver operator characteristic; BFS = Bayesian frequency selection
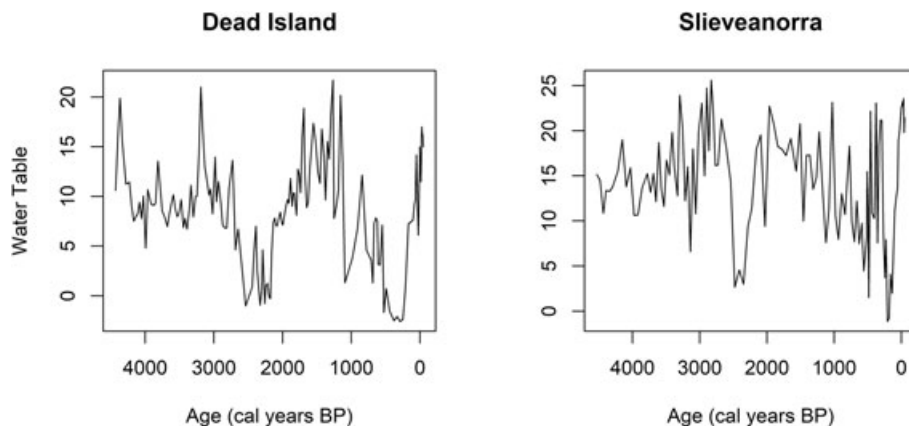
FIGURE 3 Water table records observed in Dead Island and Slieveanorra in Northern Ireland

# 5 | FREQUENCY IDENTIFICATION OF WATER TABLE RECORDS FROM MID- TO LATE HOLOCENE

In this section, we contrast results obtained via REDFIT with those of the BFS method using an example analysis of Holocene water table records extracted from two closely located ombrotrophic peatlands in Northern Ireland—Dead Island and Slieveanorra, representing lowland and upland perspectives, respectively (Swindles et al. 2012). A plot of the data is shown in Figure 3.

Prior to conducting frequency analysis, we use Bchron (Haslett & Parnell, 2008) to construct a chronology for the sites in question, utilizing the radiocarbon dates and their associated depths to produce estimated ages at all depths at which the proxy is measured. The resulting chronologies for the two cores are shown in Figure 4. Note that in the original paper (Swindles et al. 2012), a different chronology construction was used.
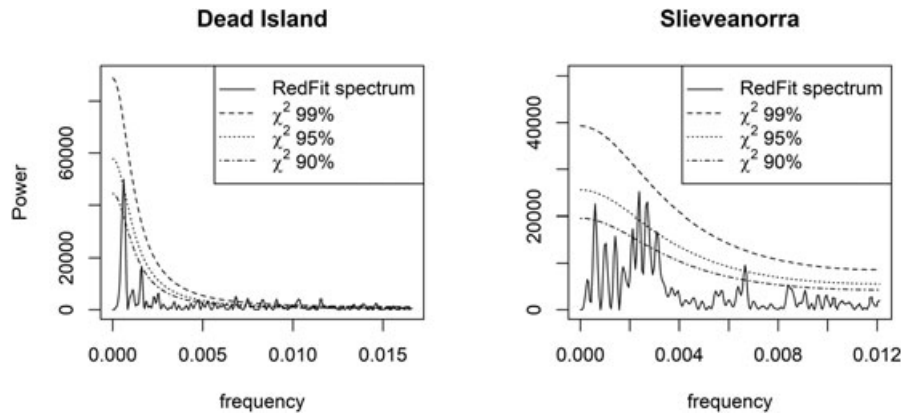
5.1 | REDFIT

We detrend the records from Slieveanorra using a 3rd degree polynomial regression and keep records from Dead Island unchanged, as there appears to be no clear trend. We apply the default setting of the REDFIT version 38e utilizing the rectangular window and set the parameter *ofac* = 4, where typical values of the ofac factor are ranging from 2 to 4 as per Schulz and Mudelsee (2010). A plot of the periodogram is shown in Figure 5, with significant levels 90%, 95%, and 99%.

At Dead Island, the REDFIT spectrum suggests 12 significant frequencies at the 99% level existing above the frequency $f = 0.0068$. At 95% and 90%, it identifies 21 and 25 frequencies, respectively. The significance level suggested by Thompson (1990) corresponds to 98.7%. At Slieveanorra, the REDFIT spectrum at the 99% level does not detect any significant frequencies. At 95% and 90% significance levels, it identifies 4 and 5 significant frequencies, respectively. The
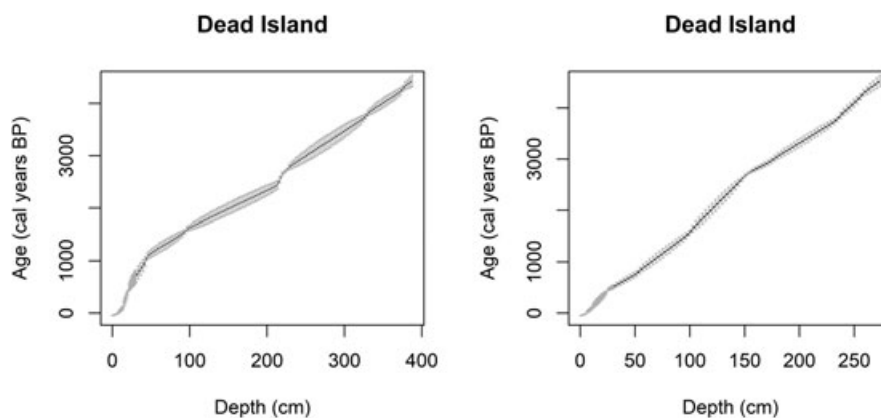
FIGURE 4 Uncertainty surrounding the time estimates using Bchron. Gray bars represent one sigma (σ_{t_i}) errors surrounding the expected timing of observations μ_{t_i}

Dead Island   ## Slieveanorra

FIGURE 5   REDFIT: Spectra of water table records observed in Dead Island and Slieveanorra in Northern Ireland

significance level suggested by Thompson (1990) corresponded to 98.2%. We compare the frequencies chosen by different methods in more detail in Section 5.3.

We note that detection of spectral peaks within the REDFIT framework may depend upon the chosen window function. We summarize the results of applying other window functions available in the REDFIT package to the records presented in our case studies in Appendix A.

5.2 | Bayesian frequency selection

We conduct the BFS analysis on the two cores. We use a cubic trend for both sites, although no detrending is required, as this is incorporated as part of the model fitting process. We use the exact same prior distributions as for the simulated example of the previous section. However, we set the maximum number of frequencies to give $k_{\text{DeadIsland}} = 75$ and $k_{\text{Slieveanorra}} = 56$ over the Nyquist range. We use JAGS to obtain posterior samples and check convergence as previously described.

The periodogram for the BFS method applied to Dead Island is shown in Figure 6. The model selects three clear frequencies with large power estimates at $f_1 = 0.00045$, $f_2 = 0.00067$, and $f_3 = 0.0016$ with the probability of indicators
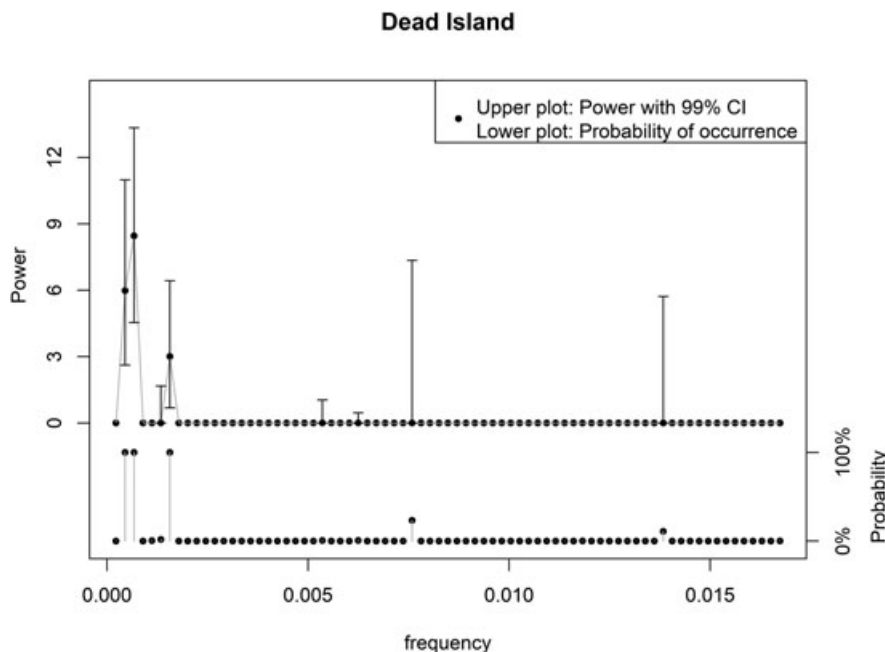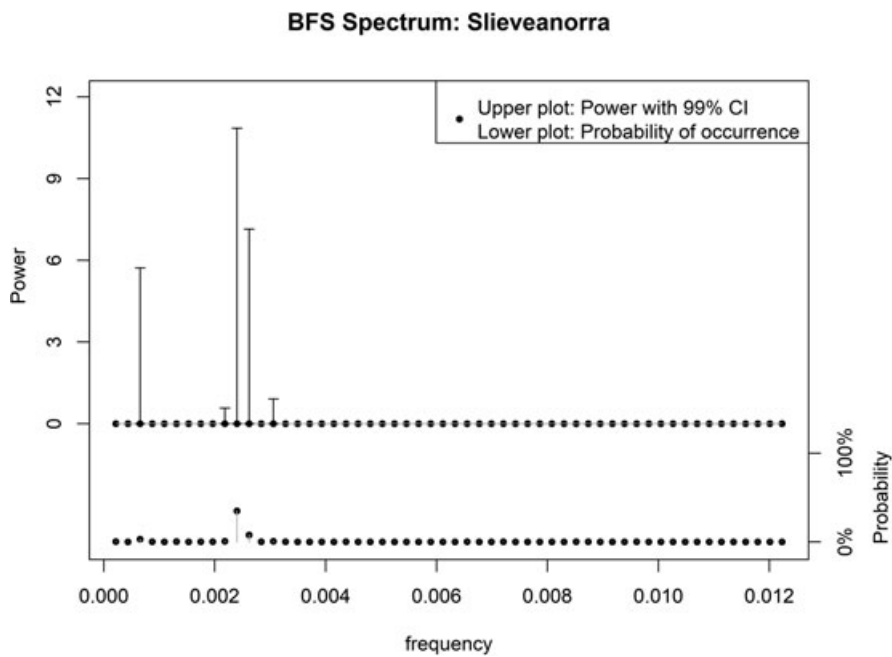
Dead Island

FIGURE 6   BFS: Dead Island periodogram with confidence intervals at 99% in the upper part, and probability indicator plot on the lower part. The BFS method identifies three significant frequencies approaching 100% and two less important frequencies with 24% and 10% of inclusion, respectively. BFS = Bayesian frequency selection; CI = Confidence interval

approaching 100% each. The equivalents in years are 2,241, 1,494, and 640 year cycles. It also selects two less important frequencies $f_4 = 0.00759$ and $f_5 = 0.00138$ with median power equal to zero, and wide confidence intervals and the probability of indicators equal to 24% and 10%, respectively. The equivalents in years are 132 and 72 year cycles. Figure 7 shows 5-fold cross-validation performance for the Dead Island model. The out-of-sample performance seems to indicate a good fit of the model to the data.

The periodogram of the Slieveanorra water table record is shown in Figure 8. The method selected two relatively unimportant frequencies $f_1 = 0.0024$ and $f_2 = 0.0026$ being the equivalent of 416 and 381 year cycles, respectively, with the
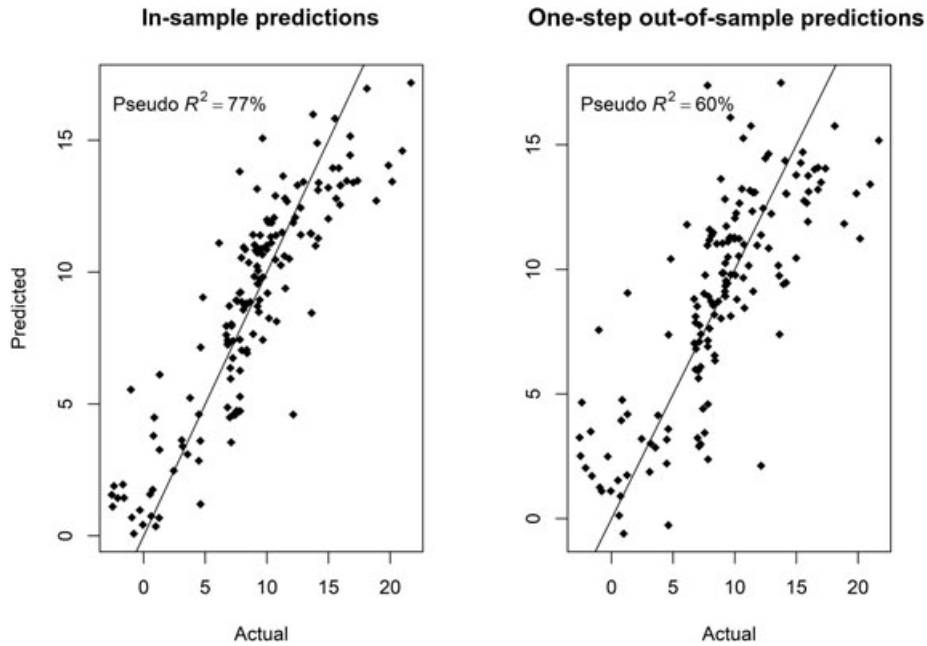
FIGURE 7   BFS: Dead Island 5-fold cross-validation shows good fit of the model, given the two sources of uncertainties, and good predictive one-step out-of-sample performance. The estimated value of the pseudo $R^2 = 0.60$ confirms a good fit of the model and also reflects the substantial amount of noise in the data. BFS = Bayesian frequency selection

FIGURE 8   BFS: Slieveanorra periodogram with confidence intervals at 99% in the upper part, and probability indicator plot in the lower part. Frequencies selected by the method are unimportant. BFS = Bayesian frequency selection; CI = Confidence interval

In-sample predictions　　　　**One-step out-of-sample predictions**
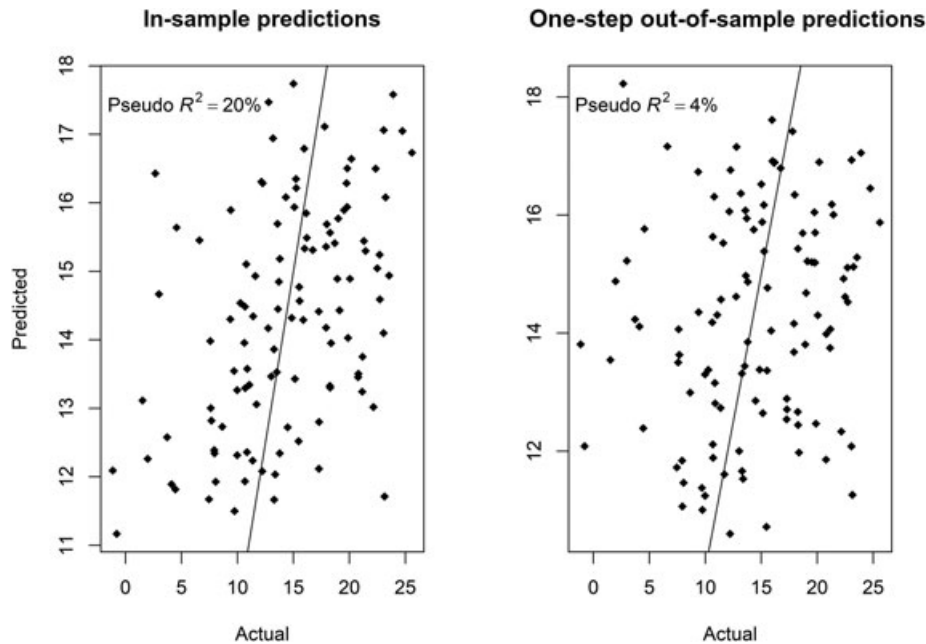
FIGURE 9 BFS: Slieveanorra 5-fold cross-validation. This shows a relatively poor fit of the model to the in-sample and out-of-sample predictions

median power estimates of $P_{f_1} = P_{f_2} = 0$ and wide confidence intervals. The probabilities of those frequencies are estimated at 38% and 8%, so there is substantial doubt as to the importance of those signals. The 5-fold cross-validation is shown in Figure 9, which is noticeably worse than that of Dead Island and perhaps indicates a lack of detectable frequency behavior for this site.

5.3 | Comparison of selected frequencies

The frequencies selected by the REDFIT and the BFS methods are largely inconsistent for the water table records in Dead Island and Slieveanorra. The inconsistency might be expected for Slieveanorra because neither method indicates strong frequency behavior (Table 1); at Dead Island, the BFS method picks up more lower frequency behavior, whereas REDFIT identifies many more higher frequency terms (Table 2). In particular, the REDFIT version we use does not take into account time uncertainty, which may be the root cause of the difference in selected frequencies. We note that Swindles et al. (2012) found the same issues when comparing REDFIT with Lomb–Scargle.

TABLE 1 Significant frequencies chosen by REDFIT versus high probability frequencies chosen by BFS for the Slieveanorra site

SI cycle	REDFIT 90%	REDFIT 95%	REDFIT 99%	BFS probability
1,681	✓	–	–	–
474	✓	–	–	–
416	✓	✓	–	34%
381	✓	✓	–	7%
369	✓	✓	–	–
325	✓	✓	–	–
150	✓	✓	–	–
119	✓	–	–	–

Note. BFS = Bayesian frequency selection; SI = Slieveanorra.

TABLE 2 Significant frequencies chosen by REDFIT versus high probability frequencies chosen by BFS for the Dead Island site

Dead Island cycle	REDFIT			BFS probability
	90%	95%	99%	
2,241	–	–	–	100%
1,630	✓	✓	–	–
1,494	–	–	–	100%
640	–	–	–	100%
212	✓	–	–	–
190	✓	✓	–	–
177	✓	–	–	–
147	✓	✓	✓	–
132	✓	✓	✓	24%
129	✓	–	–	–
120	✓	✓	✓	–
116	✓	✓	–	–
115	✓	✓	–	–
110	✓	✓	✓	–
101	✓	✓	✓	–
97	✓	✓	✓	–
90	✓	✓	–	–
86	✓	✓	✓	–
83	✓	✓	–	–
81	✓	✓	–	–
77	✓	✓	✓	–
75	✓	✓	✓	–
72	✓	✓	✓	10%
70	✓	–	–	–
68	✓	✓	✓	–
65	✓	✓	–	–
64	✓	✓	✓	–
63	✓	✓	–	–

Note. BFS = Bayesian frequency selection.

6 | CONCLUSIONS

Our BFS method introduces a novel approach to the problem of frequency identification in time-uncertain series and in series with known timing of observations. In contrast with the traditional methods of frequency analysis, the BFS model allows us to incorporate output from a chronology model, facilitating more realistic and accurate modeling of the uncertainties in the data. As we have shown, this can have a strong impact on the conclusions drawn from analysis in the frequency domain. In our simulated examples, we demonstrated that treating uncertain timing of observations as exact, and therefore neglecting the uncertainty attribute, may result in misidentification of true frequency patterns.

Aside from dealing with time uncertainty, the Bayesian framework allows us to answer direct questions of interest, such as "what is the probability that a particular frequency exists in the time series given the current data?". It allows us to move away from null hypothesis significance testing where, given a large enough data set, all frequencies will be selected as significant. Furthermore, given full access to a posterior distribution, we can create posterior uncertainty intervals, for example, for a periodogram, or any other desired quantity.

For simulated data with uncertain timings, the BFS method seems to perform better than traditional methods. It seems to be able to identify the correct frequencies more often and seems to be less prone to overfitting. Most interesting, though, is the comparison of real data sets where, at least for Dead Island, the results from BFS conflict strongly with that of

REDFIT. It seems appropriate that, at the very least, all such methods should provide out-of-sample performance metrics for each site, such as that we provide, for example, in Figure 9.

The BFS model, being a standard Bayesian model, allows for many possibilities for extension. It would be reasonably simple to apply the model in a generalized linear modeling framework, for example, for pollen counts that might be Poisson distributed. Other avenues for further research include the possibility of incorporating the chronology model within the BFS model, so as jointly to estimate sedimentation rates and frequency behavior. A broader challenge would be to extend to dynamic frequencies under time uncertainty, although identifying and validating models here will be exceedingly tricky.

REFERENCES

Bergmeir, C., & Benitez, J. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–216.

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*, 1257–1266.

Bowman, S. (1990). *Radiocarbon dating*. Berkeley, CA: University of California Press.

Brooks, S., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Dey, D., Ghosh, S., & Mallick, B. (2000). *Generalized linear models: A Bayesian perspective*. Boca Raton, FL: CRC Press.

Essefi, E., Mefteh, S., Medhioub, M., & Yaich, C. (2014). Magnetic study of the heated and unheated sedimentary fillings of Sebkha Mhabeul, Southeast Tunisia: A geophysical method for paleoclimatic investigation and tephrochronological dating. *International Journal of Geophysics*, *2014*, 1–7.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–511.

George, E., & McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*, 881–889.

George, E., & McCulloch, R. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, *7*, 339–373.

Gupta, H., Mehra, R., & Batan, S. (2013). Power spectrum estimation using Welch method for various window techniques. *International Journal of Scientific Research and Engineering Trends*, *2*, 389–392.

Haslett, J., & Parnell, A. (2008). A simple monotone process with application to radiocarbon-dated depth chronologies. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *57*, 399–418.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. New York, NY: Springer.

Hays, J., Imbrie, J., & Shackleton, N. (1976). Variations in the Earth's orbit: Pacemaker of the ice ages. *Science*, *194*, 1121–1132.

Kuo, L., & Mallick, B. (1988). Variable selection for regression models. *Sankhya: The Indian Journal of Statistics, Series B*, *60*, 65–81.

Lomb, N. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, *39*, 447–462.

Mudelsee, M. (2002). TAUEST: A computer program for estimating persistence in unevenly spaced weather/climate time series. *Computers & Geosciences*, *28*, 69–72.

Mudelsee, M. (2009). *Climate time series analysis, classical statistical and bootstrap methods*. New York, NY: Springer.

Mudelsee, M., Scholz, D., Rothlisberger, R., Fleitmann, D., Mangini, A., & Wolff, E. (2009). Climate spectrum estimation in the presence of timescale errors. *Nonlinear Processes in Geophysics*, *16*, 43–56.

Parnell, A., & Gehrels, R. (2015). Using chronological models in late Holocene sea level reconstructions from salt marsh sediments. *Handbook of sea level research*. Chichester, Western Sussex, UK: Wiley.

Plummer, M. (2017). JAGS version 4.3.0 user manual [Computer software manual]. Retrieved from sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/

Rhines, A., & Huybers, P. (2011). Estimation of spectral power laws in time uncertain series of data with application to the Greenland Ice Sheet Project 2 δ^{18}O record. *Journal of Geophysical Research: Atmospheres*, *116*, 1–9.

Scargle, J. (1982). Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal*, *263*, 835–853.

Scargle, J. (1989). Studies in astronomical time series analysis. III. Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. *Astrophysical Journal*, *343*, 874–887.

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

Schulz, M., & Mudelsee, M. (2010). REDFIT 3.8e. Estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. *28*, 421–246. Retrieved from www.geo.uni-bremen.de/geomod/staff/mschulz/software/redfit38e.zip

Siegel, A. (1980). Testing for periodicity in a time-series. *Journal of the American Statistical Association*, *75*, 345–348.

Swets, J. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics*. New Jersey: Lawrence Erlbaum Associates.

Swindles, G., Patterson, T., Roe, H., & Galloway, J. (2012). Evaluating periodicities in peat-based climate proxy records. *Quaternary Science Reviews*, *41*, 94–103.

Thompson, D. (1990). Time series analysis of Holocene climate data. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *330*, 601–616.

Trauth, M. (2015). *MATLAB recipes for earth sciences*. New York, NY: Springer.

Vaníček, P. (1969). Approximate spectral analysis by least-squares fit. Successive spectral analysis. *Astrophysics and Space Science*, *4*, 387–391.

Warner, N., & Domack, E. (2002). Millennial-to decadal-scale paleoenvironmental change during the Holocene in the Palmer Deep, Antarctica, as recorded by particle size analysis. *Paleoceanography*, *17*, PAL 5-1–PAL 5-14.

APPENDIX A

INFLUENCE OF APPLYING DIFFERENT WINDOW FUNCTIONS ON FREQUENCY SELECTION IN THE REDFIT FRAMEWORK

As noted in Section 2.3, the choice of a window function within the REDFIT framework may influence the results of hypothesis testing. We extend the REDFIT analysis presented in our case study of the water table records to Dead Island and Slieveanorra in Northern Ireland (Section 5) and apply the remaining window functions available in this package (Figures A1 and A2). We summarize the results by contrasting counts of significant peaks above 90%, 95%, and 99% credible intervals in Table A1.
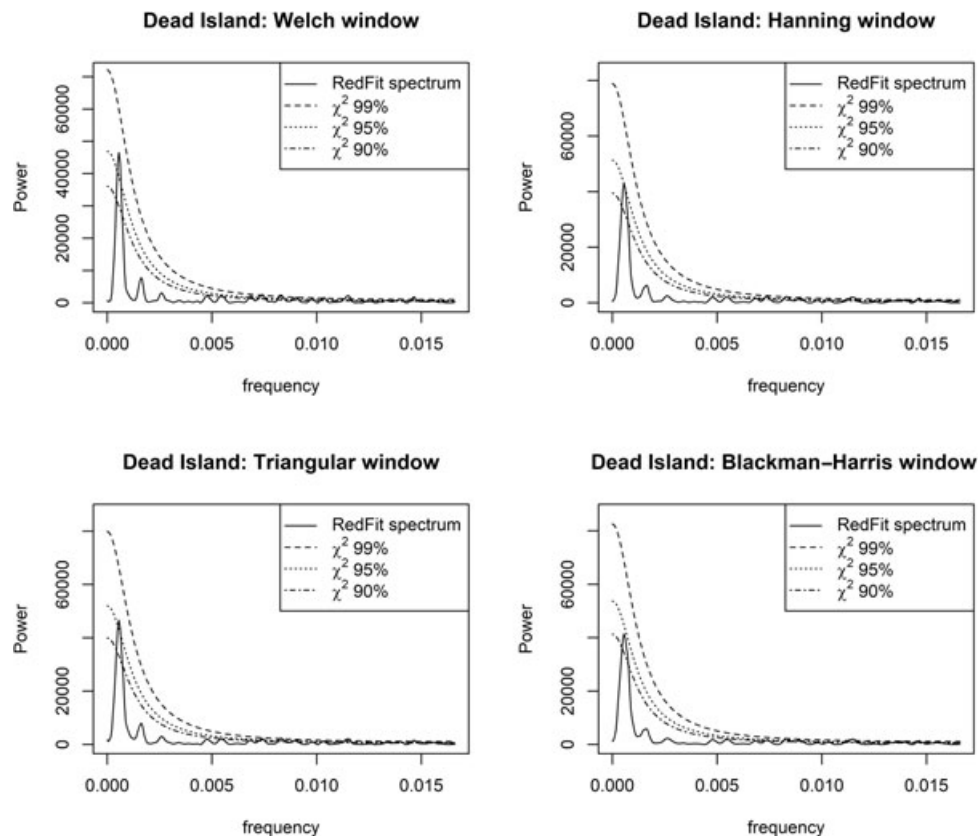
FIGURE A1 REDFIT: Spectra of water table records observed in Dead Island in Northern Ireland generated using four different window functions
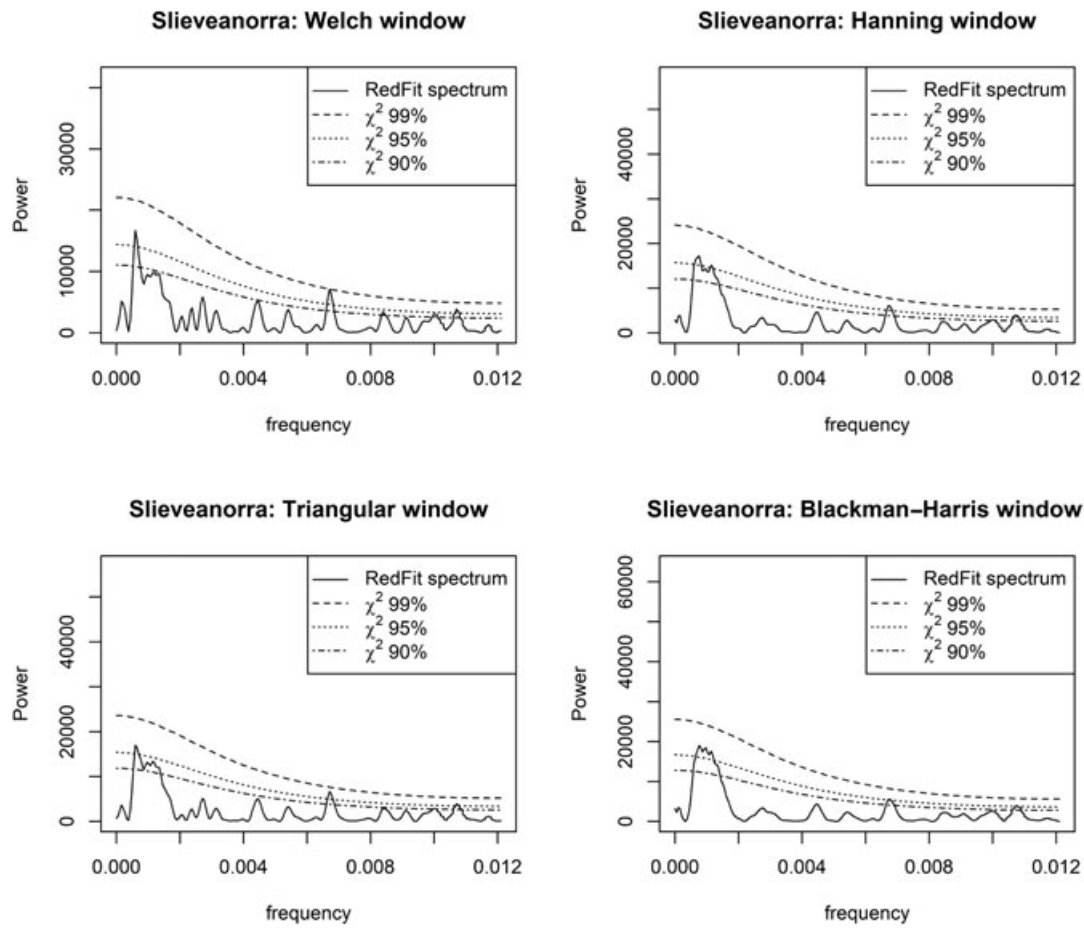
FIGURE A2 REDFIT: Spectra of water table records observed in Slieveanorra in Northern Ireland generated using four different window functions

TABLE A1 Summary of significant frequencies selected by the REDFIT method using different window functions

Window	Dead Island			Slieveanorra		
	90%	95%	99%	90%	95%	99%
Welch	13	12	6	5	3	–
Hanning	11	9	3	5	4	–
Triangular	12	10	3	4	3	–
Blackman–Harris	9	7	2	3	3	–
Rectangular	25	21	12	7	4	–

Note. Results show lack of consistency in the case of the records from Dead Island, and consistency in the case of the records from Slieveanorra

The obtained results are inconsistent. It is hard to choose a cutoff point or window method that works well. We were also unable to find any wider research determining the choice of a window function in the domain of time uncertainty. We noticed that the Welch window is generally preferred (Gupta, Mehra, & Batan 2013) in the domain of evenly spaced data.

APPENDIX B

BFS: THE JAGS CODE

We present the JAGS code of our implementation of the BFS model used in this research.

```
model {

 for(i in 1 : n) {

        y[i] ~ dnorm(mu[i], precision)

        mu[i] <- const + alpha1 * ti.sim.st[i] + alpha2 * ti.sim.st[i]^2 +

                alpha3 * ti.sim.st[i]^3 + inprod(X[i,], IBeta)

        }

 for(j in 1 : (2 * n.frequencies)) {

        beta[j] ~ dnorm(0, 0.01)

        }

 for(j in 1 : n.frequencies) {

        M[j] ~ dunif(0, m)

        Ind[j] ~ dbern(M[j] / n.frequencies)

        IBeta[j] <- Ind[j] * beta[j]

        Ind[j + n.frequencies] <- Ind[j]

        IBeta[j + n.frequencies] <- Ind[j + n.frequencies] *

                        beta[j + n.frequencies]

        }

 for(i in 1 : n) {

        for(j in 1:n.frequencies) {

                X[i,j] <- sin(2 * pi * ti.sim[i] * frequencies[j])

                }

        for(j in (n.frequencies + 1) : (2 * n.frequencies)) {

                X[i,j] <- cos(2 * pi * ti.sim[i] *

                        frequencies[j - n.frequencies])

                }

        }

 for (i in 1 : n) {

        ti.sim.tmp[i] ~ dnorm(ti.mu[i], 1 / ti.sd[i]^2)
```

```
        }
ti.sim <- sort(ti.sim.tmp)

ti.sim.st <- (ti.sim-ti.mu.mu)/ti.sd.mu

const ~ dnorm(const.mean, const.prec)

precision ~ dgamma(0.01, 0.01)

alpha1 ~ dnorm(0, 0.01)

alpha2 ~ dnorm(0, 0.01)

alpha3 ~ dnorm(0, 0.01)}
```

where

- *const* is the constant value;
- *alpha*1, *alpha*2, and *alpha*3 are the coefficients of polynomial regression;
- *beta* is the vector of Fourier coefficients of cosine and sine terms;
- *Ind* is the vector of indicator variables;
- *IBeta* is the vector of products of Fourier coefficients and indicator variables;
- *X* is the matrix of sine and cosine terms;
- *ti.mu* is the vector of estimated timing of observations;
- *ti.mu.mu* is the mean of *ti.mu*;
- *ti.sd* is the vector of standard deviations of timings of observations;
- *ti.sd.mu* is the mean of standard deviations of timings of observations;
- *ti.sim.tmp* is the vector of simulated estimates of the timing of observations;
- *ti.sim* is truncated *ti.sim.tmp*;
- *ti.sim.st* is standardized *ti.sim.tmp*;
- *m* is the expected maximum number of frequencies in the data;
- *frequencies* is the vector of frequencies; and
- *n.frequencies* is the number of frequencies.