

# Maximum entropy spectral analysis of uneven time series in the geosciences

Eulogio Pardo-Igúzquiza<sup>(1)</sup> and Peter A. Dowd<sup>(2)</sup>

(1) Instituto Geológico y Minero de España (IGME),  
Ríos Rosas 23, 28001 Madrid, Spain.  
e.pardo@igme.es

(2) Faculty of Engineering, Computer and Mathematical Sciences,  
The University of Adelaide, Australia.  
peter.dowd@adelaide.edu.au

## ABSTRACT

Time series with a non-constant sampling interval (i.e., uneven time series) are ubiquitous in the geosciences. This is due to random sampling, gaps in sampling, missing data, hiatuses, or the transformation between a spatial scale and a temporal scale when, for example, the sedimentation rate is not constant. The preferred approach in the spectral analyses of these uneven sequences are interpolation-free spectral methods, and the Lomb-Scargle periodogram is a popular choice. In the work presented here, the maximum entropy spectral estimator, modified to deal with uneven time series, is proposed as an alternative to the periodogram. The appeal of this approach is that the maximum entropy spectral estimator is a high resolution estimator. The proposed methodology uses the equivalence between the maximum entropy and the autoregressive spectral estimator. The permutation test is used to assess the statistical confidence of the estimated power spectrum and real and simulated time series are used to illustrate the performance of the proposed methodology. This study shows that the maximum entropy spectral estimation of uneven time series avoids the side lobe problem that plagues the Lomb-Scargle periodogram whilst maintaining its high resolution for short time series. The maximum spectral estimator works well in cases where large proportions of the data are randomly missing, when there are gaps in the data, where there is one or more significant hiatus in the process that generates the data, and for time series with random sampling.

Keywords: maximum entropy, autoregressive, missing data, hiatus, permutation test.

## ***Análisis espectral de máxima entropía para series irregulares en geociencias***

### RESUMEN

*Las series temporales con un intervalo de muestreo no constante aparecen con mucha frecuencia en Ciencias de la Tierra. Esto se debe a que el muestreo ha podido ser aleatorio, o porque ha habido fallos en la toma de datos, hiatos, o por la transformación de una escala espacial (sondeo, columna estratigráfica, ...) a una escala temporal cuando las tasas de sedimentación no son constantes. El método preferido para el análisis espectral de estas series temporales con muestreo irregular son los métodos que no requieren de una interpolación explícita de la serie para pasar a un muestreo regular, sino que tratan directamente los datos con espaciado no constante y, de entre estos métodos, el periodograma de Lomb-Scargle ha sido una elección muy frecuente por parte de los investigadores. En el trabajo aquí presentado, el estimador de máxima entropía, modificado para tratar con series temporales con muestreo no constante, se propone como una alternativa al periodograma. El atractivo de la metodología que se propone es que el estimador de máxima entropía es un estimador de alta resolución. El método propuesto aprovecha la equivalencia que existe entre el estimador espectral de máxima entropía y el estimador espectral autorregresivo. Además, se utiliza el test de permutación para evaluar el nivel de confianza estadístico de espectro de potencia estimado. Asimismo, se utilizan series temporales reales y simuladas para ilustrar el resultado de aplicar la metodología propuesta. Este trabajo muestra que el estimador de máxima entropía para series temporales irregulares evita los problemas de los lóbulos laterales que plagan las estimaciones por el periodograma de Lomb-Scargle mientras que se mantiene la alta resolución del estimador espectral de máxima entropía para series temporales cortas. Los resultados con el estimador de máxima entropía han resultado ser satisfactorios en los casos en los que importantes proporciones de los datos se han perdido al azar, cuando hay huecos a intervalos regulares en la serie de datos, cuando hay uno o más hiatos en la serie así como para series donde el muestreo ha sido aleatorio.*

*Palabras clave: máxima entropía, autorregresivo, datos perdidos, hiato, test de permutación.*

## Introduction

Spectral analysis of time series is extensively used in seismology (Park et al., 1987), paleoclimatology (Pestiaux and Burger, 1984) and cyclo-stratigraphy (Schwarzacher, 1975, 1993; Weedon, 2003). One of the problems that has been resolved by spectral analysis is detecting periodic signals that are hidden in noisy time series. The procedure consists of the discrete sampling of a variable of interest in time or in space to generate an experimental time series to which any of the many available spectral estimators can be applied (Marple, 1987). In a second step, the statistical significance of the estimated power spectrum is assessed, and in a final step, the physical significance of the statistically significant spectral peaks is interpreted. In this manner, for example, periodic components that have been found in stratigraphic sequences (Schwarzacher, 1975) and other proxies of paleoclimatic data have been linked to Milankovitch cycles related to variations in the Earth's insolation caused by changes in its orbital parameters (Schwarzacher, 1993). There are many approaches that have been proposed to estimate the power spectrum of even time series, i.e., time series for which the data have been regularly sampled with a constant sampling ratio, providing a constant distance between consecutive data. Examples of computer programs for power spectrum estimation with regular sequences are given in Pardo-Igúzquiza et al. (1994) and Pardo-Igúzquiza and Rodríguez-Tovar (2004, 2005) for the periodogram, the Blackman-Tukey spectral estimator, the Thomson multi-taper in signal processing, and maximum entropy estimators.

Uneven time series are common in the geosciences. There are many reasons for this including hiatuses or gaps in stratigraphic sequences (Weedon et al., 2019), gaps in time series (i.e., when, for various reasons, measurements during certain periods are not valid), and the general missing-data problem where data are missing at random locations in the complete data set. In geoscience applications, uneven sampling is very frequent when data are measured in a spatial sequence, for example along a borehole or a stratigraphic succession. Even if a constant sampling interval is used, the resulting time series will be uneven if the sedimentation rate is not constant along the succession.

For time series with a constant sampling interval and only a few randomly missing data, the missing values should be able to be reliably estimated by some form of interpolation which enables the spectral analysis of the reconstructed sequence to produce reliable estimates. Although there is a potential problem in defining "few" in this context, it is generally accepted

to be less than 10% of the complete sequence (Pardo-Igúzquiza and Rodríguez-Tovar, 2013). For larger amounts of missing data, or when samples are randomly located (random sampling), it is preferable to avoid interpolation because it modifies the spectral content of the sequence (Schulz and Stettegger, 1997; Schulz and Mudelsee, 2002). Some spectral methods have been adapted to particular situations such as periodically gapped data (Larsson and Li, 2003). Deconvolution has been proposed as a means of removing artefacts introduced by missing data (Roberts et al, 1987; Heslop and Dekkers, 2002) and, theoretically, most spectral estimators can be adapted to deal with uneven time series (e.g., Stoica and Sandgren, 2006; Babu and Stoica, 2010). In geoscience applications, the Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982) is the most popular choice for the spectral analysis of uneven time series, perhaps because of the ample choice of public domain computer programs for conducting the analysis, for example the computer codes in Press et al. (1992), Schulz and Stettegger (1997), Schulz and Mudelsee (2002) and Pardo-Igúzquiza and Rodríguez-Tovar (2012), among others. Notwithstanding the popularity of this method, it suffers from the problems attributed to the periodogram estimator such as the non-decreasing estimation variance (Zhang et al., 2005) which is addressed by smoothing the estimates, and the side lobes inherent in the estimation that can be mitigated by deconvolution (the CLEAN procedure, Heslop and Dekkers, 2002). The maximum entropy (ME) spectral estimator is a high resolution spectral estimator that performs well even for short time series (Ables, 1974; Egozcue, 1980; Pardo-Igúzquiza and Rodríguez-Tovar, 2006). Some drawbacks of the maximum entropy estimator have been reviewed in Pardo-Igúzquiza and Rodríguez-Tovar (2006).

## Methodology

The maximum entropy spectral estimator (Burg, 1967, 1975; Papoulis, 1984) is the spectral power spectrum that maximizes the entropy ( $E$ ):

$$E = \int_{-\pi}^{\pi} \ln\{S(\omega)\} d\omega, \quad (1)$$

over the class of all power spectra that satisfy the constraints given by making (theoretical) auto-covariances equal to sample auto-covariances (i.e., auto-covariances estimated from the data):

$$\int_{-\pi}^{\pi} e^{i\omega h} S(\omega) d\omega = \hat{C}(h), \quad (2)$$

for  $h \in [-q, q + 1, \dots, -1, 0, 1, \dots, q]$ ,

where  $\hat{C}(h) = \hat{C}(-h)$  is the estimated covariance for lag  $h$ ,  $i$  is the imaginary unit and  $\omega = 2\pi f$  is the angular frequency (radians per sampling interval);  $f$  is the frequency in cycles per sampling interval.

Brockwell and Davis (1991) show that the ME spectral estimator has the form:

$$\hat{S}(\omega) = \frac{\sigma_M^2}{|1 + \sum_{k=1}^M a_k e^{-ik\omega}|^2} \quad (3)$$

where  $\sigma_M^2$  is the variance of a zero-mean white noise stochastic process that depends on the order of the AR process, and  $\{a_k; k=1, \dots, M\}$  are the  $M$  autoregressive coefficients that also depend on the chosen order  $M$ . These  $M + 1$  parameters must be estimated from the experimental data in order to apply the ME spectral estimator of Eq. (3). One way of estimating these parameters is to use the Yule-Walker equations (Papoulis, 1984). To avoid the inversion of large covariance matrices, the Durbin-Levinson algorithm (Brockwell and Davis, 1991) can be used to estimate recursively the AR coefficients and noise variance. The spectral estimator in equation (3) is equal to the autoregressive (AR) spectral estimator of order  $M$  (Ulrych and Bishop, 1975). Strictly speaking, the AR estimator in Eq. (3) is the ME estimator when the random function (stochastic process) is Gaussian (Papoulis, 1984).

A requirement for a maximum entropy spectral estimator for uneven time series is that it must yield estimates that are identical to those of the AR spectral estimator. The equivalence of the two estimators can be demonstrated for any data set by using the maximum entropy spectral estimator implemented in MAXENPER (Pardo-Igúzquiza and Rodríguez-Tovar, 2005) and the autoregressive spectral estimator implemented in AUTORSE2 (Pardo-Igúzquiza *et al.*, 2000). Broersen *et al.* (2004a) used autoregressive spectral analysis for cases when observations are missing and estimated the AR models using an approximate maximum likelihood function (Bos *et al.*, 2002; Broersen *et al.*, 2004b). A more straightforward procedure is considered here. A zero-mean time series of  $N$  data:

$$\{z(t_1), z(t_2), z(t_3), \dots, z(t_N)\} \quad (4)$$

where:

$$t_i = (i - 1)\Delta + t_1, \quad (5)$$

can be considered a realization of a second-order stationary random process.

The time series in equation (4) provides a complete, even sequence in the interval  $[t_1, t_N]$  with constant sampling ratio  $\Delta$ . The auto-covariance function (Papoulis, 1984):

$$\hat{C}(k) = \frac{1}{N} \sum_{i=1}^{N-k} (z(t_i)z(t_{i+k})), \quad k = 1, \dots, M \quad (6)$$

completely determines the parameters of the AR process which, in turn, completely determine the power spectrum estimates. The parameter  $M$  is the number of terms in the autocorrelation process used for estimating the power spectrum. The main difficulty in applying the AR spectral estimator is the choice of an optimal value of this parameter and various methods have been proposed for choosing it. Some methods are based on criteria such as the Akaike information criterion or the final error prediction (Ulrych and Bishop, 1975). The order is often chosen empirically with a value between  $N/5$  and  $N/2$ , where  $N$  is the number of experimental data points.  $M = 2N/\ln(2N)$  is another popular empirical choice. If the chosen order is small, the variance of the estimates will be small, but the bias will be large, whereas if the chosen order is large, the bias will be small, but the variance will increase.

An alternative estimator of the autocovariance is obtained by replacing the factor in equation (6) by the factor which provides an unbiased estimator. However, this can produce inconsistent estimated power spectra, such as power spectrum estimates with negative values (Priestley, 1981; Chatfield, 1991).

If the time series is uneven but the data are on a regular grid (i.e., sampling with a constant interval) and there are missing data, irrespective of whether there are hiatuses, randomly missing data, gapped data or samples randomly located on a regular grid, the autocovariance function can still be estimated using equation (6) by introducing a flag value to identify locations that have missing data and not include those terms in the estimation of the covariance in equation (6). Alternatively, these terms can be set to zero by assigning zero values at the missing locations. The new estimator is:

$$\hat{C}^*(k) = \frac{1}{N} \sum_{i=1}^{N-k} (z^*(t_i)z^*(t_{i+k})), \quad k = 1, \dots, M \quad (7)$$

$$z^*(t_j) = \begin{cases} z(t_j) & t_j = \text{location with data} \\ 0 & t_j = \text{missing data location} \end{cases} \quad j = 1, \dots, N \quad (8)$$

If the sampling is completely random then the sampling locations  $\{t_1 < t_2 < t_3 < \dots < t_{(N-1)} < t_N\}$  are random in the interval  $[t_1, t_2]$ . In this case, there are two possibilities for dealing with the problem. The first method is to assign the random experimental locations to locations on a regular grid and set as missing data the locations of the regular grid that do not have assigned data. In this case the autocovariance function can be estimated using equation (6). The second method is to approximate the estimator of the covariance functions on a regular grid of suitable auto-covariance lags. In this case, a lag, a lag tolerance and the number of lags ( $M$ ) must be defined, and the power spectrum is estimated in the same manner using the estimated auto-covariance.

A well-known, difficult problem, even for complete data sets, is establishing the confidence level of the estimated power spectrum using the maximum entropy estimator for uneven time series (Baggeroer, 1976). The process is very laborious and strong assumptions are required to obtain analytical equations that assess the reliability of the maximum entropy spectral estimates (Burshtein and Weinstein, 1987; Percival and Walden, 1993). In this case, it is more efficient to use a Monte Carlo approach that uses computationally intensive methods such as the permutation test (Efron and Tibshirani, 1993; Good, 2000; Pardo-Igúzquiza and Rodríguez-Tovar, 2000). In this methodology the power spectrum is calculated for the original series as well as for a large number of random permutations of the original time series. A permuted time series is defined for the same original locations and the experimental data of the original time series are assigned at random to the different experimental locations. When data are missing, only the locations with data are included in the random permutation. Thus, by randomly permutating the original data, it is very likely that any signal or cyclic pattern that was present in the original data will be destroyed and the null hypothesis is that the random permutations are different realizations of pure white noise. The maximum entropy power spectrum is calculated for each permuted time series and for a large number of permutations, for example 1,000 as a minimum. The achieved significance level (ASL) can be calculated as:

$$ASL(f_j) = \frac{1}{L} \sum_{i=1}^S I_i(f_j), \quad (9)$$

$$I_i(f_j) = \begin{cases} 1, & S(f_j) \leq S_i^*(f_j) \\ 0, & S(f_j) > S_i^*(f_j) \end{cases} \quad (10)$$

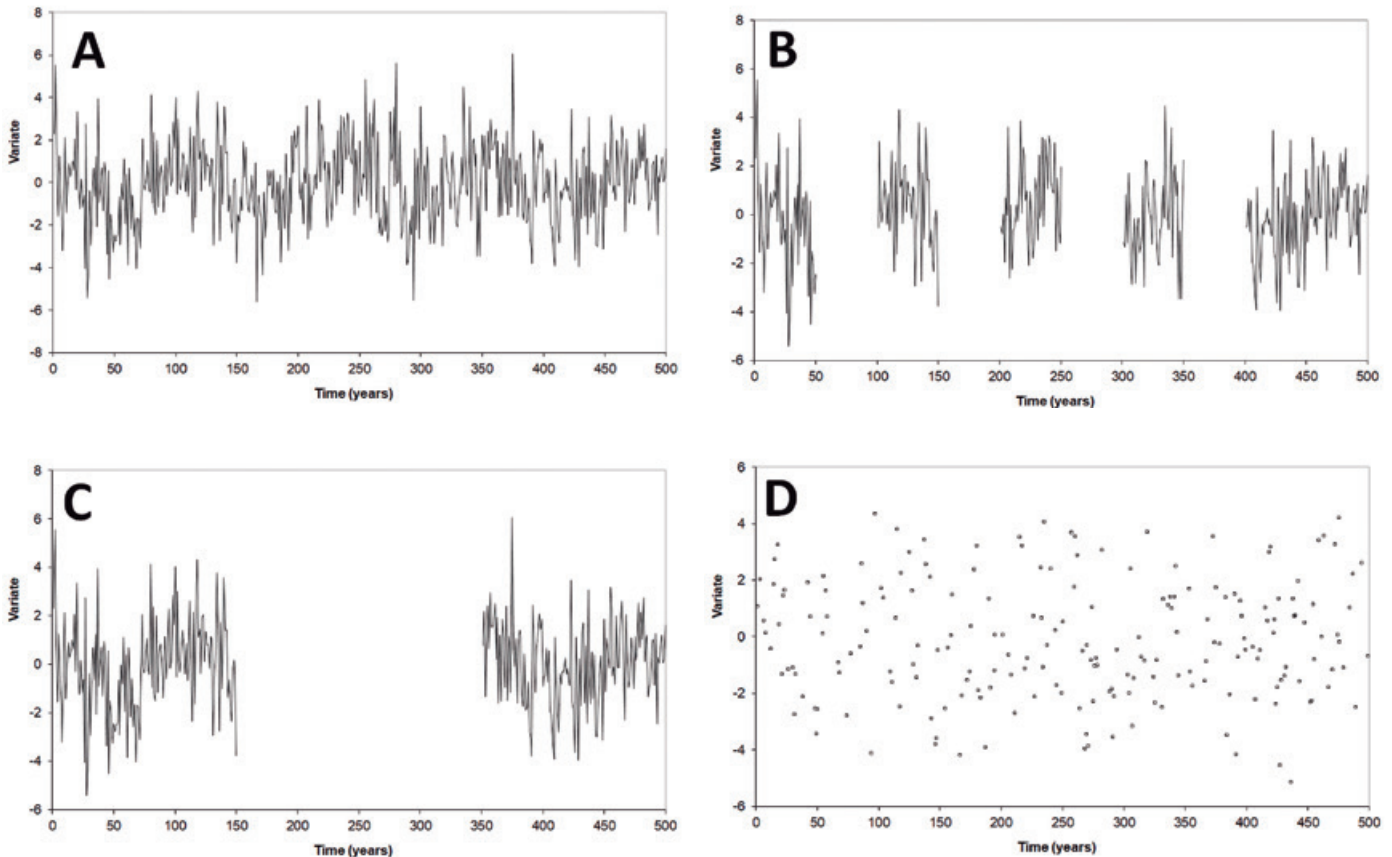
where  $S(f)$  is the ME power spectrum estimated from the original sequence and for frequency  $f_j$ ;  $S_i^*(f_j)$  is the ME power spectrum estimated from the  $i^{th}$  randomly permuted sequence, for the same frequency. The achieved confidence level is then calculated, as a percentage

$$ACL(f_j) = 100[1 - ASL(f_j)]. \quad (11)$$

## Results and discussion

### Case studies with simulated data sets

Figure 1A shows an even time series that was simulated with a signal and the addition of zero mean Gaussian random noise and variance equal to 3. This time series is complete (i.e., even and with no missing data) with 500 data and a constant sampling ratio of year. The signal comprises three cyclic components (sinusoids) with equal amplitudes of one, phases of  $45^\circ$ ,  $22.5^\circ$  and  $15^\circ$  degrees respectively and frequencies of 0.008, 0.05 and 0.4 respectively. The corresponding periods are 125 years, 20 years and 2.5 years. The variance of the signal is 1.5 and the variance of the noise is 3, giving a signal to noise ratio of 0.5. The total variance of the time series in Figure 1 is 4.2 and thus each cyclic component explains 12% of the total variance. The maximum entropy (ME) or autoregressive (AR) spectral estimate of this complete sequence is shown in Figure 2A in which the spectral peaks of the power spectrum have been correctly identified for this case of an even and complete time series with 500 data. It should be noted that, in maximum entropy spectral estimation, the variance of each cyclic component is proportional to the area under the curve of the estimated power spectrum and not to the height of the spectral peak. Three cases of irregular sampling that illustrate gapped data, central hiatuses and random sampling have been obtained and are shown in Figures 1B, 1C and 1D respectively. Figure 1B is an example of gapped data, that is, a time series with hiatuses at regular intervals. This time series has been obtained from the time series in Figure 1A and has a total of 300 data, that is, there are 200 data missing from the complete sequence. Figure 1C shows a time series with a single large hiatus in its central section. It was obtained from the complete sequence in Figure 1B by deleting 200 points and retaining the remaining 300 data. Finally, Figure 1D represents an irregular time series with 200 data obtained by randomly selecting 200 data from the 500 data of the original time series. Note that the data are still on the regular grid de-



**Figure 1.** A. Complete and even time series with 500 data. B. Gapped time series obtained from the complete time series. The number of data is 300. C. Time series with a hiatus in the central part and obtained from the complete time series. The number of data is 300. D. Time series obtained from the complete time series by selecting 200 samples at random. Note that this is not a case of pure random sampling but random sampling from the 500 samples available from the complete time series. In the latter case the points have not been represented by lines whereas in the first three (A to C) the points have not been represented but the lines that join them.

**Figura 1.** A. Serie completa y con intervalo de muestreo constante, esto es, serie regular, con 500 datos. B. Serie temporal con huecos a espaciado regular obtenida de la serie completa. El número de datos es de 300. C. Series temporales con un hiato en la parte central y obtenida a partir de la serie completa. El número de datos es de 300. D. Serie temporal obtenida a partir de la serie completa por la elección de 200 datos al azar. Hay que hacer notar que no se trata de un muestreo aleatorio puro, sino un muestreo aleatorio de entre las 500 muestras disponibles a intervalo de muestreo constante en la serie completa. En el último caso los puntos no se han unido por líneas mientras que en los tres primeros casos (A a C) los puntos no se han representado pero se han representado las líneas que los unen.

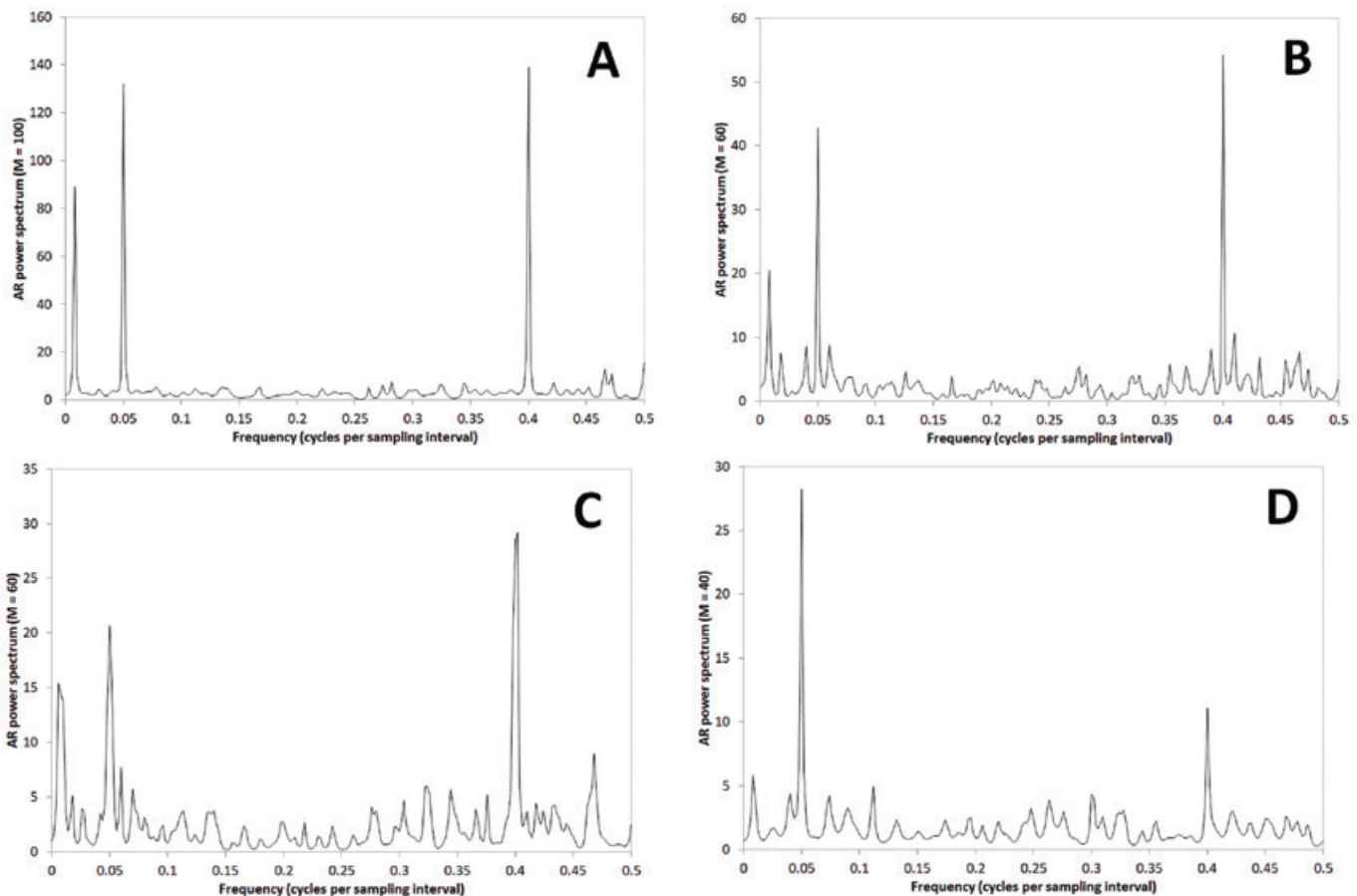
defined by the locations of the complete sequence. This is a simpler case than purely random sampling in which the locations are selected at random from an interval. Thus, in the random time series in Figure 1D there are no data in the complete sequence that are closer than the constant sampling ratio of one year. In purely random sampling two consecutive data could be closer than one year. The ME power spectrum estimates are shown in Figures 2B, 2C and 2D for gapped, central hiatus and random time series respectively. In these figures the spectral peaks are clearly identifiable in all cases although, as shown by the achieved significance level in Figure 3, for the gapped data time series (Figure 1B), only three periodicities are identified with a high confidence level (> 99%). For the time series with the

large central hiatus (Figure 1C), three periodic components have been identified with a high confidence level (> 99%) but there is a spurious spectral peak in the high frequencies of the power spectrum in Figure 2C, which has been identified with a high confidence level (> 99%) but not as high as the true cyclic components that have confidence levels higher than 99.9%. Finally, in the random sampling, two of the three components have been identified with a high level of confidence (> 99%) but the low frequency periodicity with the spectral peak in Figure 2D, has a confidence level of only 79.25%. A comparison of the maximum entropy spectral estimation of uneven time series (Figures 2 and 3) with the results of the Lomb-Scargle periodogram estimation of these same time series in Figures 4, 6 and 8

of Pardo-Igúzquiza and Rodríguez-Tovar (2013), clearly shows that the advantages of the former are that it does not generate spurious spectral peaks that appear as side lobes in the power spectrum and its 99% statistical confidence in the permutation test.

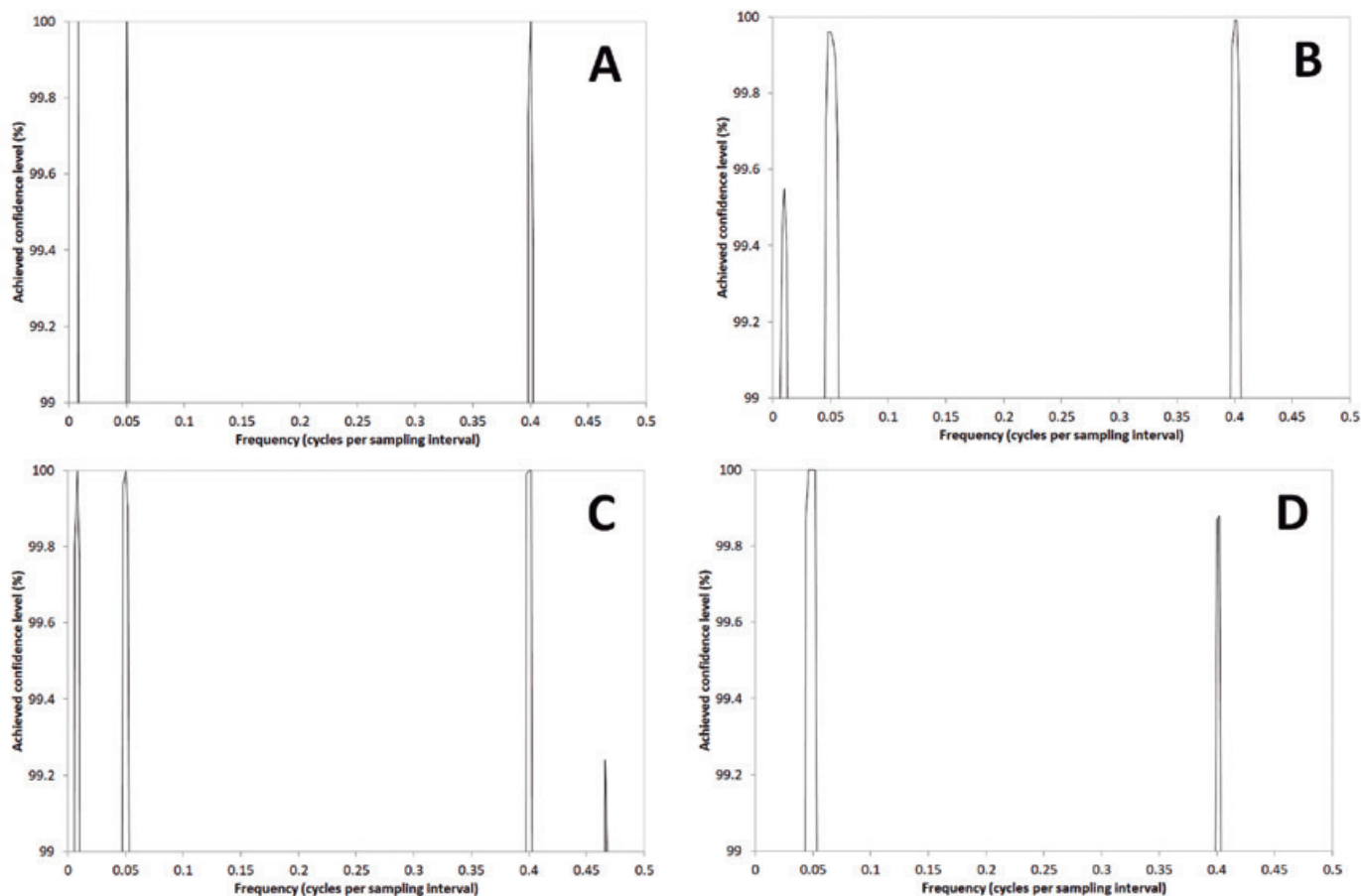
The second experiment is the estimation of the power spectrum of the sequence of 100 simulated data in Press et al. (1992, upper figure 13.8.1, p. 571) and represented in Figure 4A. This is a noisy time series that has a hidden periodicity frequency of 0.81 cycles per arbitrary time unit (period  $1/0.81 = 1.2345$  time units). We use grid locations with a constant sampling interval of 0.1 units and assign the experimental data in Figure 4A to the closest experimental location as shown in Figure 4B. The achieved significance level is shown in Figure 4C. The spectral peak at a frequency of 0.81 (period  $1/0.81 = 1.2345$ ) is clearly identified in Figure 4B and, using the permutation test and 10,000

random permutations of the original data, the estimated significance level is higher than 99.8%. The alternative approach was to apply the maximum entropy spectral estimator by estimating the auto-covariance using an auto-covariance lag of 0.2 and a lag tolerance of 0.1. The complete set of  $N(N-1)/2$  auto-covariance pairs is calculated and assigned to the appropriate auto-covariance intervals. The resulting estimated power spectrum is shown in Figure 5A and the achieved significance level using the permutation test is shown in Figure 5B. Although the target frequency of 0.81 cycles per time unit is also identified, its statistical significance is less than the 99.8% level that was achieved by the first alternative. Furthermore, a harmonic of the main frequency has been identified as statistically significant ( $> 99\%$ ) as can be seen in Figure 4C. Thus, the first procedure is preferable for the maximum entropy spectral estimator with uneven time series and purely random sampling. Note that



**Figure 2.** Maximum entropy (ME) or autoregressive (AR) spectral estimates of the time series shown in Figure 1, from A to D respectively for the complete (A), gapped (B), series with central hiatus (C) and random locations (D). The number of autoregressive terms are 100, 60, 60 and 40 respectively, that is,  $N/5$ , where  $N$  is the number of experimental data in Figure 1.

**Figura 2.** Estimación espectral, por el método de máxima entropía (ME) o método autorregresivo (AR), de la serie temporal mostrada en la Figura 1 y de A a D, respectivamente, para la serie completa (A), serie con huecos a espaciado regular (B), serie con un hiato central (C) y serie con muestreo aleatorio (D). El número de términos del proceso autorregresivo considerado es de 100, 60, 60 y 40 respectivamente, esto es,  $N/5$ , donde  $N$  es el número de datos experimentales de la serie temporal mostrada en la Figura 1.



**Figure 3.** Achieved confidence levels in percentages assessed by the random permutation test using 10,000 random permutations of the ME spectral estimates of the time series shown in Figure 1, from A to D respectively for the complete (A), gapped (b), series with central hiatus (C) and random locations (D).

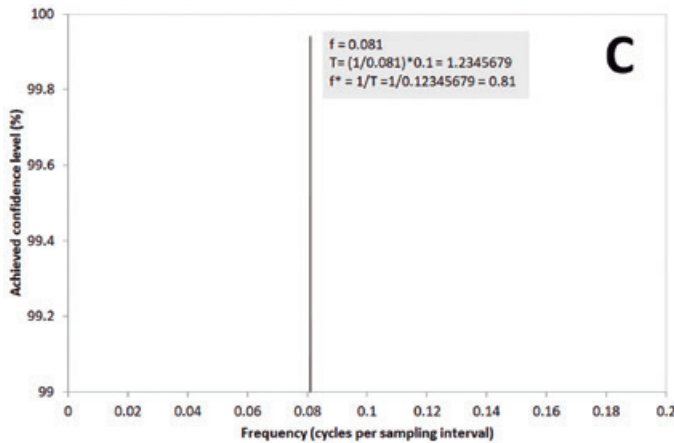
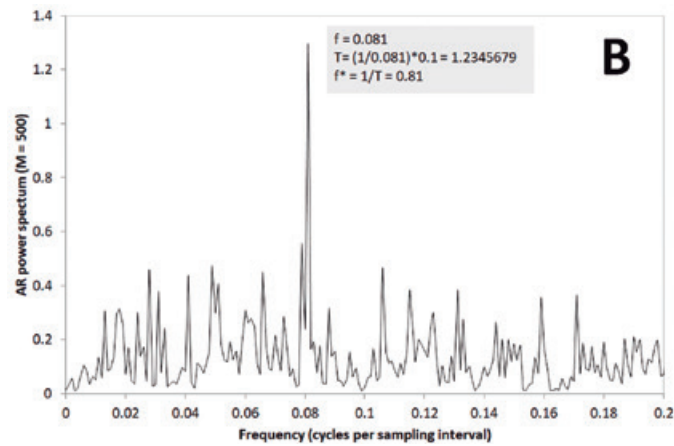
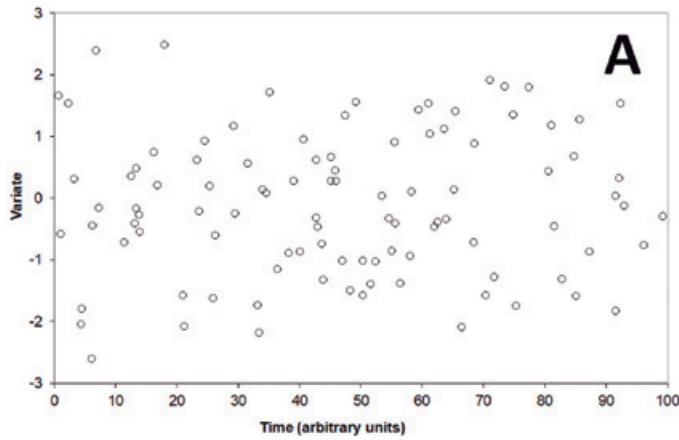
**Figura 3.** Nivel de confianza (en porcentaje) obtenido por el test de permutación con 10,000 permutaciones aleatorias, para la estimación espectral por máxima entropía de la serie temporal mostrada en la Figura 1 y considerando la serie completa (A), la serie con huecos a espaciado regulares (B), con un hiato central (C) y con muestreo aleatorio (D).

this case of purely random sampling overcomes the limitation of the Nyquist frequency that would be obtained with the same amount of data in an even time series. The 100 data in an even time series have a constant sampling interval of = 1 arbitrary time unit and frequencies of up to the Nyquist frequency of 0.5 and periodicities starting at two time units could be investigated. It would be impossible to find the hidden signal with frequency 0.81 and a period of 1.2345. This observation was also made by Press et al. (1992) for the Lomb-Scargle periodogram of this same simulated data. The Lomb-Scargle periodogram of this same time series is also presented in Pardo-Igúzquiza and Rodríguez-Tovar (2012).

**Case studies with real data**

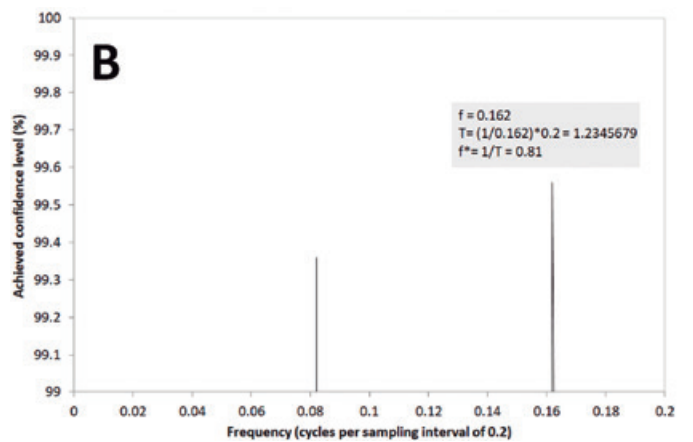
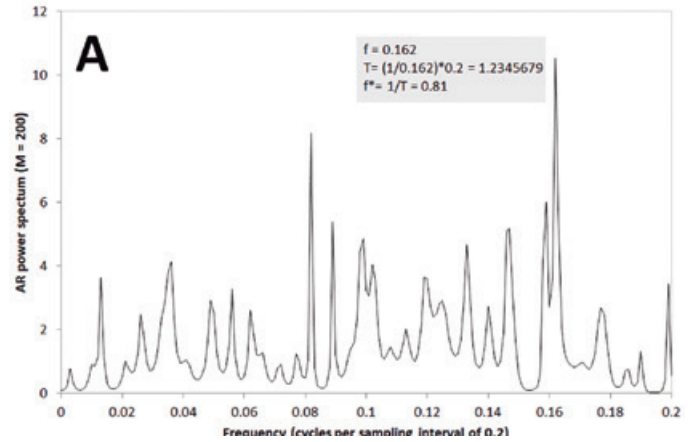
Two real data sets are shown in Figure 6. Figure 6A shows the variation in oxygen isotope O (Raymo

et al., 1992) along a core from the Deep Sea Drilling Project in the North Atlantic. The time series has 866 data with a constant sampling interval of kyrs. The maximum entropy power spectrum, estimated using the complete sequence in Figure 6A, is shown in Figure 7A and the confidence level achieved by the permutation test using 10,000 random permutations is shown in Figure 8A. The Milankovitch orbital cycles (Schwarzacher, 1993) of precession, obliquity and eccentricity are clearly identified by the spectral estimation. The very high, statistically significant spectral peaks correspond to periods of 216, 153, 124, 96, 76, 55, 41, 23.6, 22.4 and 19 kyrs and demonstrate that the sediments of the deep ocean are an excellent proxy of paleoclimatic conditions. As an experiment, the original complete sequence has been transformed to an uneven time series by deleting at random 50%, 75% and 90% of the data. The new data sets have 433, 217



**Figure 4.** A. Uneven time series identical to those in Press et al. (1992, Figure 13.8.1, p. 571) and with 100 data. B. Maximum entropy power spectrum estimate for the data given in A. C. Achieved confidence level for the estimated power spectrum in B.

**Figura 4.** A. Serie temporal irregular idéntica a la presentada en Press et al. (1992, Figura 13.8.1, p. 571) y que tiene 100 datos. B. Potencia espectral, estimada por máxima entropía, de la serie temporal mostrada en A. C. Nivel de confianza obtenido para la estimación espectral mostrada en B.

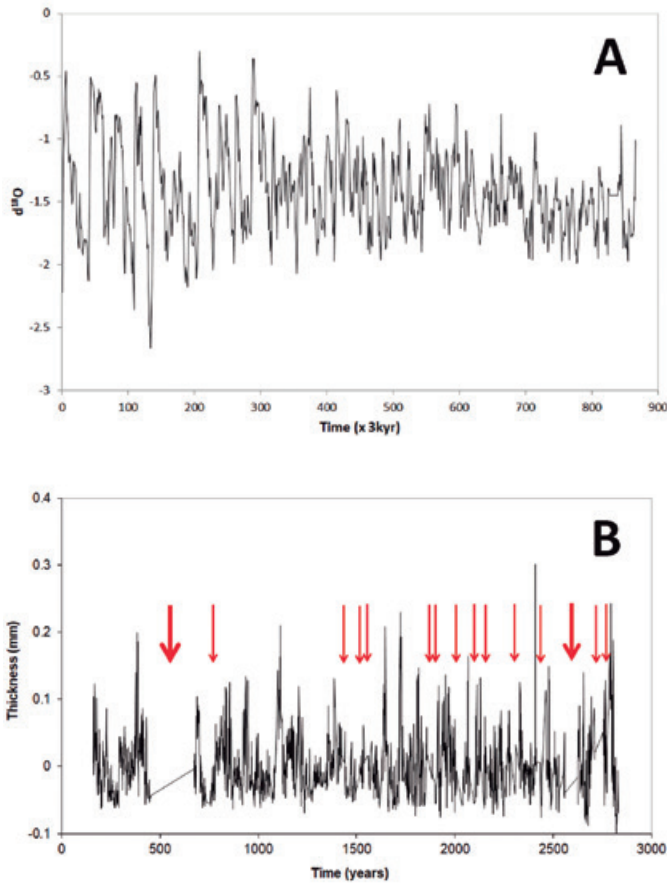


**Figure 5.** A. Maximum entropy power spectrum estimate of the data given in Figure 4A. B. Achieved confidence level for the estimated power spectrum in A.

**Figura 5.** A. Potencia espectral, estimada por máxima entropía, de los datos mostrados en la figura 4A. B. Nivel de confianza alcanzado por la estimación de la potencia espectral mostrada en A.

and 87 data respectively. Maximum entropy power spectrum estimation was applied to the new uneven sequences and the estimated power spectra are shown in Figures 7B, 7C and 7D for 50%, 75% and 90% of missing data respectively. The achieved confidence levels of the spectral estimates evaluated by the permutation test using 10,000 random permutations are shown in Figures 8B, 8C and 8D for 50%, 75% and 90% of missing data respectively. Figures 7B and 8B show that when 50% of the data are missing, the main Milankovitch spectral peaks in the precession, obliquity and eccentricity bands are still detected. The same can be said even when 75% of the data are missing, as shown in Figures 7C and 8C. Furthermore, in the extreme case, when 90% of the original data are missing, important information can still be recovered as the obliquity cycle and one of the eccentricity cycles are detected.





**Figure 6.** Real time series used in the case studies. A. Time series of oxygen isotope O (Raymo et al., 1992) for the last 2.7 million years. B. Detrended time series of thickness of the annual banding in a stalagmite in the Carlsbad cave (New Mexico, USA) for the last 3000 years before the present (Polyak and Asmerom, 2001). Some hiatuses have been highlighted by red arrows with thick red arrows marking the main ones.

**Figura 6.** Series temporales reales utilizadas en los casos de estudio. A. Serie temporal del isótopo de oxígeno O (Raymo et al., 1992) para los últimos 2.7 millones de años. B. Serie temporal, donde se ha quitado la tendencia, del espesor del bandeo anual de una estalagmita de la cueva Carlsbad (Nuevo Méjico, EEUU) para los últimos 3000 años antes del presente. (Polyak and Asmerom, 2001). Algunos hiatos se han remarcado con las flechas rojas con las de mayor grosor señalando a los hiatos más importantes.

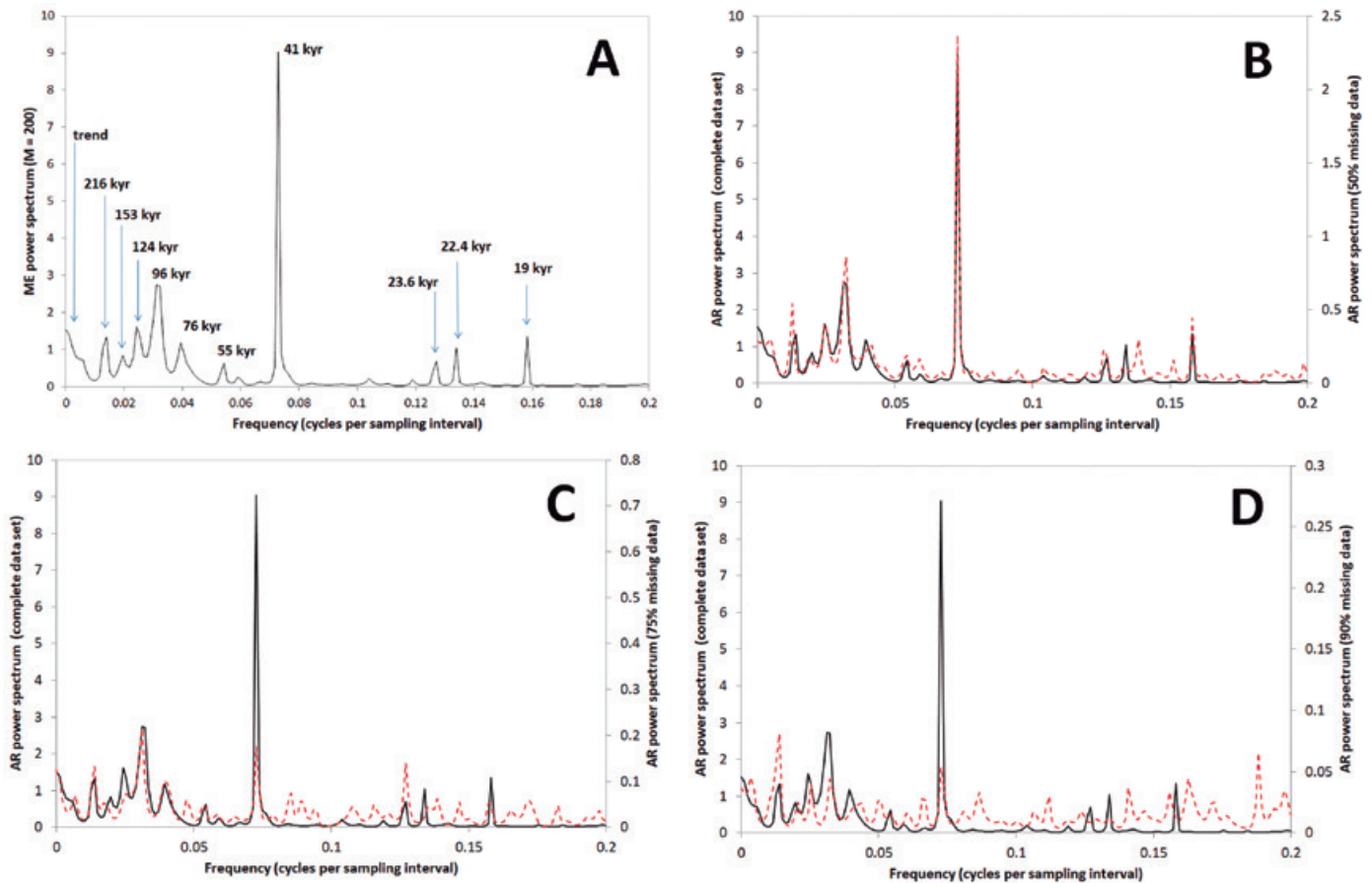
The second real time series is shown in Figure 6B and represents the detrended time series of the thickness (in mm) of the annual banding in a stalagmite in the Carlsbad cave (New Mexico, USA) over the last 3000 years before the present (Polyak and Asmerom, 2001). Some hiatuses have been highlighted by red arrows in Figure 6B. The time series has 1896 data on a regular grid with a constant sampling interval of kyr but with hiatuses. The hiatuses are denoted by a flag value of -9999, which is used in the AUTORSE5 computer program for missing data, and the final length of the series is 2672 locations that have experimental or missing data. Thus, despite

the appearance of Figure 6B, 30% of the sequence is missing data. The estimated maximum entropy power spectrum and the confidence achieved by the permutation test using 10,000 random permutations are shown in Figures 9A and 9B respectively. The order chosen for the auto-correlation model was  $M = 400$  which is approximately  $N/5$ , where  $N$  is the number of experimental data. The very high statistically significant spectral peaks correspond to periods of 243, 148, 76, 59, 51, 45, 36, 27.8, 26.2, 22.8, 13.4 and 11 yrs, showing that speleothems also contain excellent proxies of paleoclimatic conditions, in this case in relation to solar activity and its cycles that have an effect on climate. Thus the 11-year sun spots cycle (Gnevyshev, 1977), the 22-year Hale cycle (Gnevyshev and Ohl, 1948), the Gleissberg and De Vries/Suess cycles, with timescales of 60–120 yrs and 200–300 yrs, respectively (Vecchio et al., 2017) have been clearly identified.

The detrending procedure of this time series is described in Pardo-Igúzquiza and Rodríguez-Tovar (2013) where the Lomb-Scargle periodogram was used for estimating the power spectrum of this uneven time series. A comparison of Figure 16 in Pardo-Igúzquiza and Rodríguez-Tovar (2013) with Figure 9, shows the high resolution performance of the maximum entropy spectral estimator for uneven time series and its significant advantage over the Lomb-Scargle periodogram approach.

## Conclusions

Uneven time series are often found in the geosciences (e.g., cyclo-stratigraphy, paleoclimatology, seismology) for various reasons such as different forms of missing data (hiatuses, lost data, sampling failures, etc.) and by the transformation from a space scale (boreholes, stratigraphic sections, etc.) to a time scale with non-constant sedimentation rates. Maximum entropy spectral estimation provides high-resolution estimates of the power spectrum for complete and even time series. The procedure can also be adapted to provide high-resolution power spectrum estimates for uneven time series. The performance of the estimator has been demonstrated using simulated and real time series and the statistical significance of the spectral estimates has been evaluated by the permutation test. The procedure for maximum entropy spectral estimation for uneven time series has been coded in the Fortran program AUTORSE5, which is freely available on request from the authors. Maximum entropy power spectrum estimation of uneven time series can be used jointly with other spectral estimators such as the Lomb-Scargle



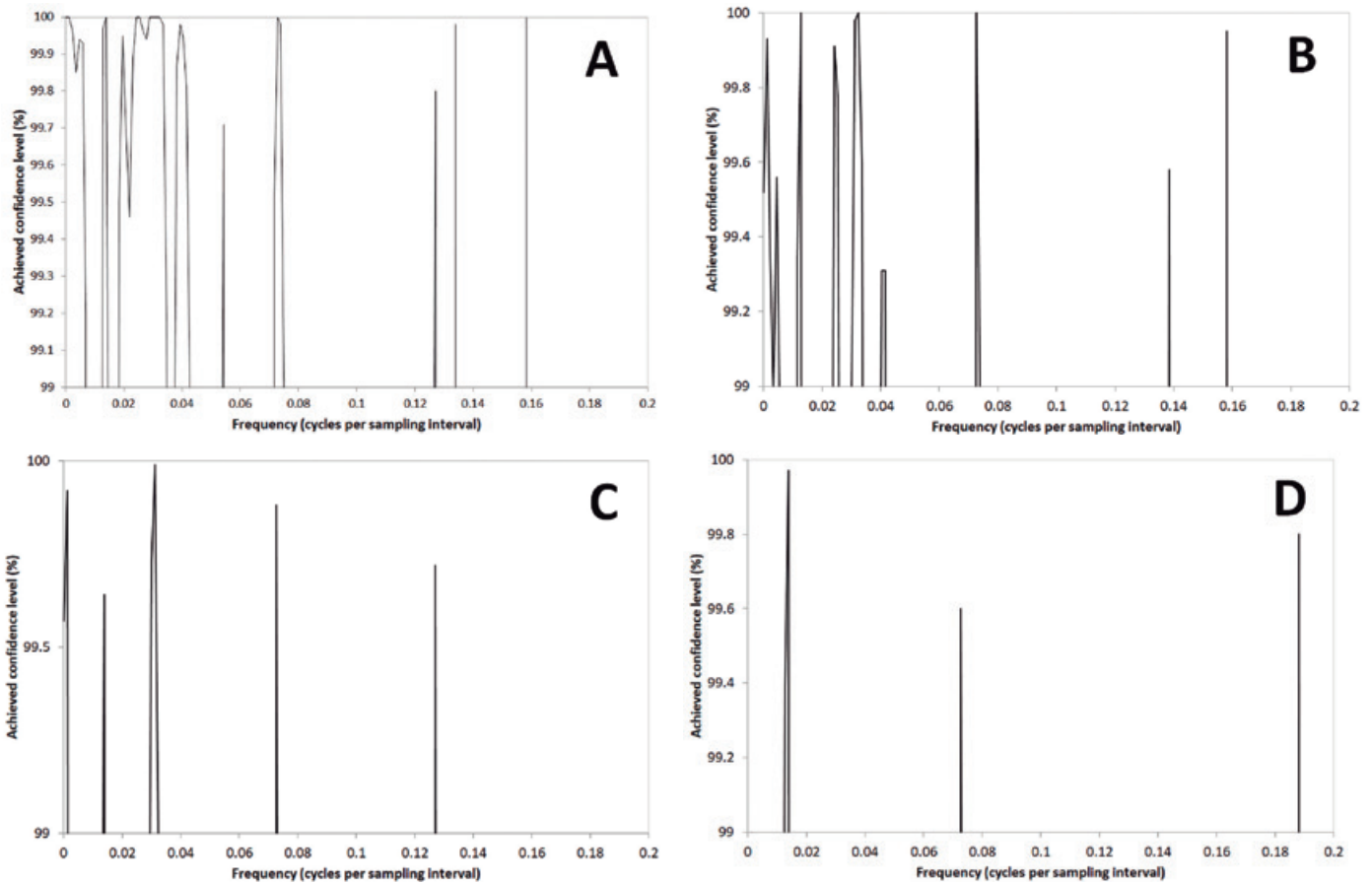
**Figure 7.** Maximum entropy spectral estimates of the data set shown in Figure 6A. A. Using the complete sequence with 866 data. B. For the sequence where 50 % of the data have been deleted at random, with 433 data remaining. C. For the sequence where 75 % of the data have been deleted at random, with 217 data remaining. D. For the sequence where 90 % of the data have been deleted at random, with 87 data remaining. The black line is the power spectrum of the complete sequence and the red dashed line represents the power spectrum of the uneven sequences.

**Figura 7.** Estimación de la potencia espectral por el método de máxima entropía de la serie temporal mostrada en la Figura 6A. A. Utilizando la secuencia regular y completa de 866 datos. B. Para la misma secuencia de la Figura 6A pero con 50% de datos perdidos, esto es con los 433 datos restantes. C. Para la misma secuencia de la Figura 6A pero con 75% de datos perdidos, esto es con los 217 datos restantes. D. Para la misma secuencia de la Figura 6A pero con 90% de datos perdidos, esto es con los 87 datos restantes. La línea negra continua es la potencia espectral obtenida con la secuencia completa y la línea roja discontinua representa la potencia espectral de las series temporales irregulares con datos perdidos.

periodogram. This is a good strategy for spectral analysis in geoscience applications as the use of several methodologies benefits from the advantages of each of them. The maximum entropy estimator has a high resolution performance with uneven time series and is free of the side lobes that cause problems in the periodogram approach.

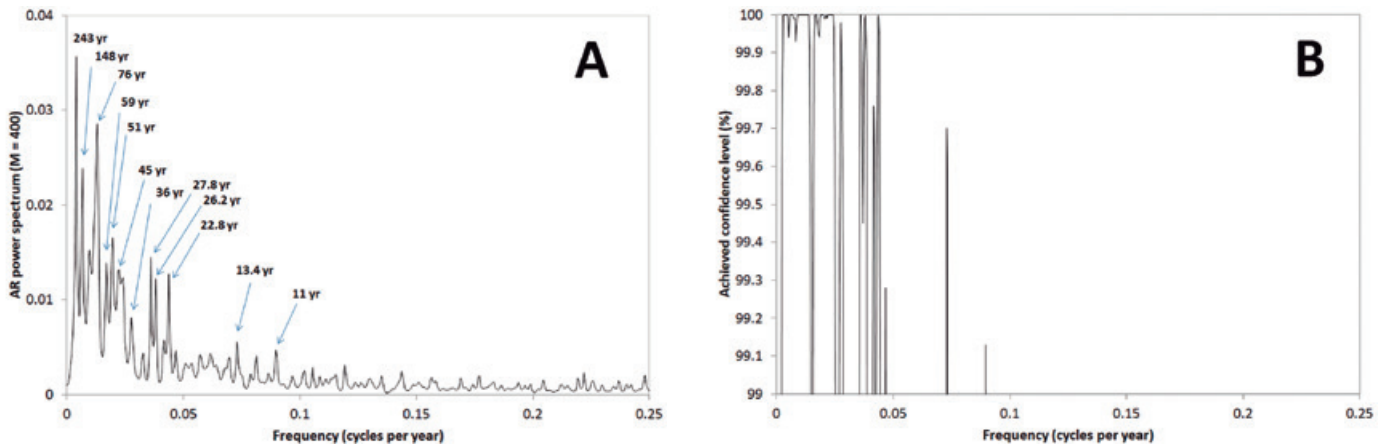
### Acknowledgements

This work was supported by research project PID2019-106435GB-I00 of the Ministerio de Ciencia e Innovación of Spain. We would like to thank the reviewers, Juan José Egózcue, Germán Rodríguez and Ricardo Olea, by their constructive comments that have helped to improve the final version of this work.



**Figure 8.** Achieved confidence level of the maximum entropy power spectrum estimates of the complete even sequence (A) and for the uneven time series with 50% of the data missing (B), 75% of data missing (C) and 90% of data missing (D).

**Figura 8.** Nivel de confianza obtenido para la potencia espectral estimada por máxima entropía de la secuencia regular y completa (A) y para las series temporales irregulares con 50% de los datos perdidos (B), 75% de los datos perdidos (C) y 90% de datos perdidos (D).



**Figure 9.** A. Maximum entropy power spectrum estimate of the uneven time series shown in Figure 7B. An autoregressive order of  $M = 400$  terms, which is around  $N/5$ , was used in the estimation. B. Confidence level obtained by the permutation test with 10,000 random permutations.

**Figura 9.** A. Potencia espectral, estimada por el método de máxima entropía, de la serie irregular mostrada en la Figura 7B. Se ha utilizado un orden de  $M = 400$  términos para el proceso autorregresivo, lo que equivale aproximadamente a  $N/5$ , siendo  $N$  el número de datos experimentales. B. Nivel de confianza obtenido por el test de permutación utilizando 10,000 permutaciones aleatorias.

## References

- Ables, J. G., 1974. Maximum entropy spectral analysis. *Astronomy and Astrophysics Supplement*, 15, 383-393.
- Babu, P. and Stoika, P. 2010. Spectral analysis of nonuniformly sampled data – a review. *Digital Signal Processing*, 20, 359-378.
- Baggeroer, A. B. (1976), Confidence intervals for regression (MEM) spectral estimates, *IEEE Transactions on Information Theory*, 22(5), 534– 545.
- Bos, R., de Waele, S. and Broersen, P.M.T., 2002. Autoregressive spectral estimation by application of Burg algorithm to irregularly sampled data. *IEEE Transactions on Instrumentation and Measurement*, 51(6), 1289-1294.
- Brockwell, P.J., Davis, R.A., 1991. *Times Series: Theory and Methods*, Second ed. Springer, New York, 577 pp.
- Broersen, P.M.T., de Waele, S. and Bos, R., 2004a. Autoregressive spectral analysis when observations are missing. *Automatica*, 40, 1495-1504.
- Broersen, P.M.T., de Waele, S. and Bos, R., 2004b. Applications of autoregressive spectral analysis to missing data problems. *IEEE Transactions on Instrumentation and Measurement*, 53 (4), 981-986.
- Burg, J.P., 1967. *Maximum entropy spectral analysis*. 37th Annual International Meeting of the Society for the Exploration of Geophysics. Oklahoma City, OK, 34–41.
- Burg, J.P., 1975. *Maximum entropy spectral analysis*. PhD Dissertation, Stanford University, 127 pp.
- Burshtein, D., and E. Weinstein (1987), Confidence intervals for the maximum entropy spectrum, *IEEE Transactions on Acoustic Speech Signal Processing*, 35, 504– 510.
- Chatfield, C., 1991, *The analysis of time series* (4th ed.). Chapman & Hall, London, 241 pp.
- Efron, B. and Tibshirani, R. J. 1993. *B. An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993, 436 pp.
- Egozcue, J. J., 1980. Una interpretación del análisis espectral de máxima entropía. *Revista de Obras Públicas*, 167-172.
- Good, P., 2000. Permutation tests. A practical guide to resampling methods for testing hypothesis. Springer, 271 p.
- Gnevyshev, M. N., 1977. Essential features of the 11-year solar cycle. *Solar Physics*, 51, 175-183.
- Gnevyshev, M. N. and Ohl, A. I., 1948. On the 22-year cycle of solar activity. *Astronomicheskii Zhurnal*, 25, 18-20.
- Heslop, D. and Dekkers, M.J. 2002, Spectral analysis of unevenly spaced climatic time series using CLEAN: signal recovery and derivation of significance levels using a Monte Carlo simulation. *Physics of the Earth and Planetary Interiors*, 130, 103-116.
- Larsson, E.G. and Li, J., 2003. Spectral analysis of periodically gapped data. *IEEE Transactions on Aerospace and Electronic Systems*, 39, 3, 1089-1097.
- Lomb, N.R. 1976. Least squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39, 447-462.
- Marple, S.L. 1987. *Digital spectral analysis with applications*. Prentice Hall. Englewood Cliffs, NY, 492 pp.
- Papoulis, A., 1984, *Probability, random variables and stochastic processes*, McGraw-Hill Intern. Editions, Singapore, 576 pp.
- Pardo-Igúzquiza, E., Chica-Olmo, M. and Rodríguez-Tovar, F.J. 1994. CYSTRATI: a computer program for spectral analysis of stratigraphic successions. *Computers & Geosciences*, 20 (4), 511-584.
- Pardo-Igúzquiza, E. and Rodríguez-Tovar, F.J. 2000. The permutation test as a non-parametric method for testing the statistical significance of power spectrum estimation in cyclostratigraphic research. *Earth and Planetary Science Letters*, 181, 175-189.
- Pardo-Igúzquiza, E., Schwarzacher, W. and Rodríguez-Tovar, F.J., 2000. A library of computer programs for assisting teaching and research in cyclostratigraphic analysis. *Computers & Geosciences*, 26, 723-740.
- Pardo-Igúzquiza, E. and Rodríguez-Tovar, F.J. 2004. POWGRAF2: a computer program for graphical spectral analysis. *Computers & Geosciences*, 30 (5), 533-542.
- Pardo-Igúzquiza, E. and Rodríguez-Tovar, F.J. 2005. MAXENPER: a program for maximum entropy spectral estimation with assessment of statistical significance by the permutation test. *Computers & Geosciences*, 31 (5), 555-567.
- Pardo-Igúzquiza, E. and Rodríguez-Tovar, F. J., 2006. Maximum entropy spectral analysis of climatic time series revisited: Assessing the statistical significance of estimated spectral peaks. *Journal of Geophysical Research: Atmospheres*, 111, D10202, 1-8.
- Pardo-Igúzquiza, E. and Rodríguez-Tovar, F.J. 2012. Spectral and cross-spectral analysis of uneven time series with the smoothed Lomb-Scargle periodogram and Monte Carlo evaluation of statistical significance. *Computers & Geosciences*, 49, 207-216.
- Pardo-Igúzquiza, E. and Rodríguez-Tovar, F.J. 2013. Análisis espectral de series temporales de variables geológicas con muestreo irregular. *Boletín Geológico y Minero*, 124 (2), 323-337.
- Park, J., Lindberg, C.R. and Vernon III, F.L. 1987. Multitaper spectral analysis of high frequency seismograms, *Journal of Geophysical Research*, 92, 12675-12684.
- Percival, D. B., and A.T. Walden (1993), *Spectral Analysis for Physical Application. Multitaper and Conventional Univariate Techniques*, Cambridge University Press, New York, 583 pp.
- Pestiaux, P. and Berger, A. 1984. An optimal approach to the spectral characteristics of deep-sea climatic records. In: Berger A. (ed.), *Milankovitch and Climate*, Part 1. D. Reidel Publishing Company, Dordrecht, 417-445.
- Polyak, V.J. and Asmerom, Y. 2001. Late Holocene Climate and Cultural Changes in the Southwestern United States. *Science*, 294, 148-151.

- Press, H. W., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. 1992. *Numerical recipes in Fortran*, (2nd ed.), Cambridge University Press, New York, 963 pp.
- Priestley, M. B. (1981). *Spectral analysis and time series*. London, U.K., Academic Press, 890 pp.
- Raymo, M. E., Hodell, D., Jansen, E., 1992. Response of deep ocean circulation to the initiation of Northern Hemisphere glaciation (3-2 myr). *Paleoceanography*, 7, 645-672.
- Roberts, D.H., Lehár, J. and Dreher, J.W. (1987) Time series analysis with CLEAN; 1: Derivation of a spectrum. *Astronomical Journal*, 93 (4), 968 – 989.
- Scargle, J.D. 1982. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal*, 263, 835-853.
- Schulz, M. and Mudelsee, M. 2002. REDFIT: estimating red noise spectra directly from unevenly spaced paleoclimatic time series. *Computer & Geosciences*, 28 (3), 421-426.
- Schulz, M. and Stattegger, K. 1997. SPECTRUM: spectral analysis of unevenly spaced paleoclimatic time series. *Computers & Geosciences*, 23 (9), 929-945.
- Schwarzacher, W., 1975. *Sedimentation Models and Quantitative Stratigraphy*. Developments in Sedimentology, Vol. 19. Elsevier, 382 pp.
- Schwarzacher, W., 1993. *Cyclostratigraphy and the Milankovitch theory*. Developments in Sedimentology, Vol. 52. Elsevier, 224 pp.
- Stoica, P. and Sandgren, N. 2006. Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches. *Digital Signal Processing*, 16, 712-734.
- Ulrych, T.J. and Bishop, T.N., 1975. Maximum entropy spectral analysis and autoregressive decomposition. *Review of Geophysics and Space Physics*, 13 (1), 183-200.
- Vecchio, A., Lepreti, F., Laurenza, M., Alberti, T. and Carbone, V., 2017. Connection between solar activity cycles and grand minima generation. *Astronomy & Astrophysics*, 559, A58, 1-12.
- Weedon, G.P. 2003. *Times-series analysis and cyclostratigraphy: examining stratigraphic records of environmental cycles*. Cambridge University Press, Cambridge, 259 pp.
- Weedon, G. P., Page, K. N. and Jenkyns, H. C., 2019. Cyclostratigraphy, stratigraphic gaps and the duration of the Hettangian Stage (Jurassic): insights from the Blue Lias Formation of southern Britain, *Geological Magazine*, 156 (9), 1469-1509.
- Zhang, Y., Baggeroer, A.B. and Bellingham, J.G., 2005. The total variance of a periodogram-based spectral estimate of a stochastic process with spectral uncertainty and its application to classifier design. *IEEE Transactions on Signal Processing*, 53 (12), 4556-4567.

Recibido: julio 2019

Revisado: noviembre 2019

Aceptado: enero 2020

Publicado: marzo 2021