

Análisis espectral de series temporales de variables geológicas con muestreo irregular

E. Pardo-Igúzquiza⁽¹⁾ y F. J. Rodríguez-Tovar⁽²⁾

(1) Instituto Geológico y Minero de España (IGME). Ríos Rosas, 23. 28003 Madrid (España)
e.pardo@igme.es

(2) Departamento de Estratigrafía y Paleontología. Avda. Fuentenueva s/n. Universidad de Granada. 18071 Granada (España)
fjrtovar@ugr.es

RESUMEN

En geociencias el muestreo más frecuente para las series temporales de variables geológicas es el muestreo irregular. Esto sucede bien porque el muestreo es aleatorio, por la ausencia parcial de datos (datos perdidos, hiatos,...), o por la transformación de la escala espacial a la escala temporal cuando la tasa de sedimentación no es constante. En todos estos casos la interpolación no es una estrategia aconsejable y es más correcto utilizar un método que trate directamente con los datos irregulares. En este trabajo se muestra cómo el periodograma de Lomb-Scargle suavizado es una elección apropiada. Se examinan los efectos del tipo de muestreo, la interpolación y la presencia de una deriva. Se propone el test de permutación como el método para calcular niveles de confianza estadística. A través de la aplicación del periodograma de Lomb-Scargle a una serie sintética se concluye la robustez del método ante los efectos anteriormente mencionados. Se ilustra la aplicación a una serie real con hiatos y que representa el espesor del bandeo anual de una estalagmita. El estudio de estos ejemplos pretende ayudar al investigador a la aplicación del periodograma de Lomb-Scargle a sus propios datos.

Palabras clave: frecuencia de Nyquist, muestreo irregular, periodograma de Lomb-Scargle, test de permutación

Spectral analysis of uneven time series of geological variables

ABSTRACT

In geosciences the sampling of a time series tends to afford uneven results, sometimes because the sampling itself is random or because of hiatuses or even completely missing data or due to difficulties involved in the conversion of data from a spatial to a time scale when the sedimentation rate was not constant. Whatever the case, the best solution does not lie in interpolation but rather in resorting to a method that deals with the irregular data. We show here how the use of the smoothed Lomb-Scargle periodogram is both a practical and efficient choice. We describe the effects on the estimated power spectrum of the type of irregular sampling, the number of data, interpolation, and the presence of drift. We propose the permutation test as being an efficient way of calculating statistical confidence levels. By applying the Lomb-Scargle periodogram to a synthetic series with a known spectral content we are able to confirm the validity of this method in the face of the difficulties mentioned above. A case study with real data, including hiatuses, representing the thickness of the annual banding in a stalagmite, is chosen to demonstrate an application using the statistical and physical interpretation of spectral peaks.

Key words: irregular sampling, Lomb-Scargle periodogram, Nyquist frequency, permutation test

ABRIDGED ENGLISH VERSION

Introduction

Spectral analysis is a statistical technique that can be used to reveal periodicities hidden in noisy time series. It has been used extensively in Earth Sciences, particularly in the geological record, where a variable

is sampled either in time or space to give a sequence, the power spectrum of which is estimated by any of the many methods available (Marple, 1987). Subsequently the statistical significance of the spectral peaks is assessed and the physical causes for the origin of the cycles can then be investigated. When the sequences are even, that is to say, the distance between two consecutive data (i.e. the sampling ratio Δ) represents the same time, standard spectral methods can be used to estimate the power spectrum, but when the sampling is uneven other strategies must be called upon, one of which is to obtain an even sequence by interpolation. Nevertheless, this is not a very satisfactory approach, it being much more accurate to use a spectral estimator that deals with the uneven sequence directly (Schulz y Stattegger, 1997). Within this context, the periodogram of Lomb-Scargle (Lomb, 1976; Scargle, 1982) proves to be an efficient choice, as we intend to demonstrate in this paper. Among the different computer programs available, the version of Pardo-Igúzquiza and Rodríguez-Tovar (2012) assesses the statistical significance of the spectral peaks by the permutation test (a Monte Carlo method that takes frequency oversampling into account (Press et al., 1992) together with the smoothing of the experimental periodogram). We have used this freely available computer program with a synthetic case study as well as with real data to show the performance of the Lomb-Scargle periodogram. For the synthetic sequence various kinds of irregular sampling were simulated to assess the effect of different types of hiatuses as well as random sampling, interpolation and the presence of a drift. In addition, we used a real sequence of annual band thickness in a stalagmite, complete with hiatuses, to show how a spectral analysis is carried out and how the results should be interpreted.

Methods

An uneven time series is defined as being a set of N experimental data $\{z(t_1), z(t_2), \dots, z(t_N)\}$ in which the time co-ordinates $t_1 < t_2 < \dots < t_N$ are arbitrary, that is, there is no constant time interval between two consecutive data. The assumption is that the experimental data are a realization of a second-order stationary random function, thus having a constant global mean and an auto-covariance that is a function of the time separation between each pair of locations only. The Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982; Press et al., 1992) is defined by Equation (1). Although the periodogram is an asymptotically unbiased estimator of the power spectrum (Chatfield, 1991), it is not consistent. Consistency is obtained by smoothing, which is the averaging out of the periodogram for neighbouring frequencies (Equation 3). There are different options for performing the smoothing (Chatfield, 1991). When the sampling is uneven the Nyquist frequency does not exist (Equation 5), although other higher frequencies can be defined as references (Equations 6 to 9). For example, the results of this paper were obtained using the computer programs described in Pardo-Igúzquiza and Rodríguez-Tovar (2012), where the highest frequency, f_M , and the number of frequencies to calculate in the interval, $(0, f_M)$, must be specified. The permutation test (Efron and Tibshinari, 1994; Good, 2000; Pardo-Igúzquiza and Rodríguez-Tovar, 2000) is used to assess the statistical significance of the spectral peaks. This is a Monte Carlo computer-intensive parametric-free procedure that takes into account both the oversampling and the smoothing of the Lomb-Scargle periodogram estimates. The significance level is calculated according to Equation (10) and the confidence level is then calculated via Equation (12).

Case studies of uneven sampling

Synthetic case study

Figure 1 shows a simulated zero-mean time series generated by adding Gaussian noise to a signal of three sinusoidal components. The parameters of the signal are $A_1 = 1$, $A_2 = 1$ and $A_3 = 1$ for the amplitudes $f_1 = 0.008$, $f_2 = 0.05$ and $f_3 = 0.4$, for the frequencies, and $\varphi_1 = 45^\circ$, $\varphi_2 = 22.5^\circ$ and $\varphi_3 = 15^\circ$ for the three phases respectively. The time series of Figure 1 contains 500 data with a constant sampling interval of 1 year. Figure 2A shows the power spectrum of the full time series in Figure 1, and Figure 2B gives the confidence level of the estimates shown in Figure 2A. All the signal components are perfectly identified when the even-time series appearing in Figure 1 is available. We, however, are interested in the effect upon the estimated power spectrum of uneven sampling. Figure 3 shows different examples of uneven sampling that were obtained from the full even time series in Figure 1. Figure 3A shows an uneven sequence of 300 data with several hiatuses during the time series. Figure 3B also contains 300 data, but with a single hiatus in the middle of the sequence. Figure 3C shows a time series with random sampling of 200 data. Figure 4A shows the Lomb-Scargle periodogram (or estimated power spectrum) of the time series in Figure 3A. It can be seen how the three components of the signal are identified but part of their spectral power leaks into the neighbouring frequencies and spurious secondary peaks are generated. The leakage is introduced because the experimental time series is finite rather than infinite. In fact, it can be seen as being the product of a sampling window and the experimental data. For example, Figure 5A is the sampling window of the experimental time series in Figure 3A. Thus the estimated periodogram turns out to be the convolution of the Fourier transform of the signal assuming an infinite time series and the Fourier transform of the window shown in Figure 5B. Because the Fourier transform of the window has a secondary lobe at frequency 0.01, as be seen in Figure 5B, there will be secondary spurious peaks

at frequencies 0.008 ± 0.01 , 0.05 ± 0.01 and 0.4 ± 0.01 , as can be seen clearly in Figure 4A. The confidence level of these spurious peaks is high but not so high as that of the signal peaks. Furthermore, the effect is predictable and interpretable in this case, where the peaks of the signal are well separated.

Figure 6A shows the Lomb-Scargle periodogram of the time series in Figure 3B. The effect of the sampling window (Figure 7A) in this case is to widen the well identified spectral peaks (Figure 7B) because the Fourier transform of the sampling window (Figure 7B) has no important secondary lobe but a wide main lobe. In this particular case the effect of a single hiatus is less pernicious than the effect of several equidistant hiatuses that add up to the same length as the single one. Figures 8A and 8B show the periodogram and the confidence level of the time series with 200 data and random sampling. The spectral peaks of the signal are well detected. Figure 3D is similar to Figure 3C but with linear interpolation to obtain even sampling. The pernicious effect of interpolation may be seen in Figures 9A and 9B, in which the high-frequency peak has not been identified and spurious peaks appear with the same confidence level as that of the signal, thus leading to an erroneous interpretation of the power spectrum. This provides a good reason why interpolation should be avoided.

Another source of error in the interpretation may be introduced by the presence of a trend. A trend has been simulated and added to the sequence of Figure 1 in order to obtain the sequence shown in Figure 3E. From the complete sequence another time series with 200 data is generated by sampling at random from Figure 3E, as shown in Figure 3F. Figure 10A shows the Lomb-Scargle periodogram of the time series in Figure 3F, and Figure 10B shows its confidence level. The worst effect is a spurious peak with a high confidence level. The trend can be estimated (Figure 11A) by a running mean and can be discounted from the time series in Figure 3F to obtain the trend-free sequence in Figure 11B. Figure 12A shows the Lomb-Scargle periodogram of the time series of Figure 11B, and Figure 12B shows its confidence level. It can be seen how successful detrending leads to a good estimation of the power spectrum of the signal as the three components are clearly identified. The lessons learned in this synthetic case study are now applied to a real case study, as described below.

Spectral analysis of a palaeoclimatic proxy in a stalagmite

Speleothems are becoming more and more important as proxies to help in the interpretation of continental palaeoclimates. As an example we have taken the annual banding of a stalagmite from Carlsbad cave in New Mexico (USA) studied by Polyak and Asmerom (2001) and conducted an independent spectral analysis upon the sequence. The time series contains 1896 data for an interval of 2671 years from 161 to 2832 years BP. The data are set out in Figure 13, in which hiatuses of a total of 775 missing bands are evident. Figure 14A shows the sampling window and Figure 14B shows its Fourier transform. It can be seen that the side lobes are not very large and that the effect of the spectral leakage is thus not expected to be very harmful. Using a running mean of 200 years, the trend is estimated (Figure 15A) and its effect discounted to obtain the detrended time series shown in Figure 15B. The estimated Lomb-Scargle periodogram is set out in Figure 16A and its confidence level appears in Figure 16B. The peaks considered to be significant represent periods of 200, 80, 48, 26.6, 22, 16 and 11 years. The 11-year cycle is the sunspot cycle and that of 22 years the magnetic sun-spot cycle. The peaks with periods of 26.9 and 78.7 years, which are very close to those of 26.5 and 80 years, have already been reported by Pardo-Igúzquiza et al. (1994) in their analysis of varves. The 16-year cycle may be similar to the peak at 16-17 years reported by Mann and Lees (1996) in relation to variations in the water level of the Great Salt lake. The 48-years cycle has also been reported in other speleothems in New Mexico and Europe. The cycle of 200 years may correspond to a long cycle of solar activity.

Conclusions

Many sequences in geosciences are difficult to sample evenly and the Lomb-Scargle periodogram constitutes an efficient way of estimating the power spectra of such sequences. The estimates are good for different kinds of irregular sampling. We have analysed the effects of the different samplings, number of samples, interpolation and detrending using a synthetic time series. The method was then applied to a similar analysis of a real time series taken from a stalagmite and the lessons learned from the artificial test used in its interpretation. The software used for the spectral analysis is in the public domain and freely available from the authors on request.

Introducción

El análisis espectral de series temporales es una técnica estadística que permite, entre otras aplicaciones, la descomposición de una serie temporal en sus componentes frecuenciales con el objetivo de descubrir

componentes cíclicas inmersas en ruido. Su uso es extensivo en varias áreas de Ciencias de la Tierra como sismología (Park *et al.*, 1987), paleoclimatología (Pestiaux y Berger, 1984) y cicloestratigrafía (Weedon, 2003). En la detección de ciclos a partir del registro geológico, la idea es muestrear una variable geológi-

ca (velocidad de ondas sísmicas, indicadores paleoclimáticos, o espesores de estratos, entre otros) y estimar su espectro de potencia a partir de la aplicación de uno de los muchos métodos existentes (Marple, 1987). Seguidamente se evalúa la significación estadística y física de los picos encontrados en el espectro de potencia estimado. La significación estadística implica verificar cuál es la probabilidad de que un ciclo de una amplitud igual a la observada pueda obtenerse por puro azar, mientras la significación física implica encontrar una causa física que pueda ser la responsable de dicho efecto de comportamiento cíclico. Son muchas las técnicas que se han diseñado para estimar el espectro de potencia cuando los datos están regularmente distribuidos, esto es, con una distancia constante entre dos datos consecutivos e igual a la razón de muestreo Δ . Ejemplos de programas de ordenador para estimar el espectro de potencia en este caso pueden encontrarse entre otros en Pardo-Igúzquiza *et al.* (1994) y Pardo-Igúzquiza y Rodríguez-Tovar (2004, 2005) para los métodos de Blackman-Tukey, periodograma, método de Thomson *multitaper* y método de máxima entropía.

Sin embargo, en geociencias, la obtención de series temporales regulares, esto es, con muestreo constante y completas, son la excepción más que la norma. Lo usual es disponer de series temporales irregulares, debido a diferentes casuísticas como pueden ser series de datos incompletas como consecuencia del muestreo aleatorio o de la presencia de hiatos, datos perdidos o series con huecos ("*gapped data*") por diferentes motivos. Asimismo un caso muy frecuente de serie irregular se produce por la inversión de una escala espacial a una escala temporal. Esto es, una variable muestreada con intervalo de muestreo constante a lo largo de un sondeo o una sucesión estratigráfica, puede dar lugar a una serie de datos con muestreo no constante a lo largo de una escala temporal debido a la existencia de tasas de sedimentación no constante.

Cuando se trabaja con una serie con datos tomados a intervalo de muestreo constante pero con unos pocos datos que se han perdido o no fueron medidos (menor al 10 % de la secuencia), se pueden imputar dichos valores por interpolación. Sin embargo, en el caso de que la ausencia de datos sea mayor o se trate de series irregulares, la interpolación modifica el contenido espectral de la secuencia y debería ser evitada (Schulz y Stattegger, 1997). En estos casos, lo más conveniente es trabajar directamente con la serie irregular.

Aunque existen diferentes alternativas a la interpolación (Heslop y Dekkers, 2002) y aunque diferentes métodos han sido analizados en revisiones de méto-

dos espectrales de series irregulares (Stoica y Sandgren, 2006; Babu y Stoica, 2010), el periodograma de Lomb-Scargle (Lomb, 1976; Scargle, 1982) es la técnica más utilizada, no solo en geociencias sino en muchas otras disciplinas. Programas de ordenador para la estimación del periodograma de Lomb-Scargle se pueden encontrar en Press *et al.* (1992), Schulz y Stattegger (1997), Schulz y Mudelsee (2002) y Pardo-Igúzquiza y Rodríguez-Tovar (2012), entre otros. En el caso de estos últimos, para la evaluación de la significación estadística de los picos de la potencia espectral se ha implementado el test de permutación. Se trata de un procedimiento de Monte Carlo que tiene en cuenta la alta correlación entre la potencia espectral estimada para frecuencias vecinas y que es introducida por el sobremuestreo de las frecuencias (Press *et al.*, 1992) y por el suavizado del periodograma que se realiza para obtener un estimador consistente. Los objetivos que se persiguen en este trabajo son la revisión del periodograma de Lomb-Scargle como estimador espectral de series irregulares y el test de permutación para evaluar la significación estadística de los picos estimados. Para ello se ha utilizado los programas informáticos de Pardo-Igúzquiza y Rodríguez-Tovar (2012) aplicados a una serie sintética donde también se simulan diferentes modos de muestreo irregular, así como a una serie irregular de datos reales. Los datos utilizados son de dominio público (disponibles a partir de los autores) y pueden utilizarse para el aprendizaje y para chequear programas informáticos de análisis espectral de datos con muestreo no constante. El periodograma de Lomb-Scargle y el test de permutación se revisan brevemente en la próxima sección.

Metodología

Una serie temporal con muestreo irregular se puede considerar como una colección de N datos experimentales $\{z(t_1), z(t_2), \dots, z(t_N)\}$ donde las coordenadas temporales $t_1 < t_2 < \dots < t_N$ son arbitrarias, esto es, sin asumir una distancia constante entre cada pareja de datos consecutivos. Se considera que los datos experimentales constituyen una realización de una función aleatoria estacionaria de segundo orden, lo que implica una esperanza matemática constante e independiente del tiempo, así como la presencia de una autocovarianza (momento de segundo orden) que solo depende del tiempo entre muestras experimentales. Es decir, tanto la auto-covarianza como la varianza de la serie son finitas.

El periodograma de Lomb-Scargle se define como (Lomb, 1976; Scargle, 1982; Press *et al.*, 1992):

$$I(\omega_j) = \frac{1}{2s^2} \left\{ \frac{\left[\sum_{i=1}^N (z(t_i) - m_z) \cos[\omega_j(t_i - \tau)] \right]^2}{\sum_{i=1}^N \cos^2[\omega_j(t_i - \tau)]} + \frac{\left[\sum_{i=1}^N (z(t_i) - m_z) \sin[\omega_j(t_i - \tau)] \right]^2}{\sum_{i=1}^N \sin^2[\omega_j(t_i - \tau)]} \right\}, \quad (1)$$

donde el parámetro τ se define como:

$$\tan[2\omega_j\tau] = \frac{\sum_{i=1}^N \sin[2\omega_j t_i]}{\sum_{i=1}^N \cos[2\omega_j t_i]}, \quad (2)$$

y donde m_z y s^2 son la media y varianza estimadas de los datos $\{z(t_1), z(t_2), \dots, z(t_N)\}$.

ω_j es la frecuencia angular (ciclos por radián), mientras $f_j = \omega_j / (2\pi)$ es la frecuencia en ciclos por intervalo de muestreo. El parámetro τ hace que el estimador $I(\omega_j)$ de la ecuación (1) sea invariante con respecto al tiempo. $I(\omega_j)$ es el periodograma de Lomb-Scargle para la frecuencia ω_j .

Aunque el periodograma es un estimador asintóticamente insesgado de la potencia espectral, su varianza no decrece conforme aumenta el número de datos N (Chatfield, 1991). Un estimador consistente se obtiene por suavizado del periodograma mediante un proceso de medias móviles que promedia valores de $2m+1$ frecuencias vecinas para obtenerse el estimador:

$$I^*(\omega_j) = \sum_{k=j-m}^{j+m} \lambda_k I(\omega_k), \quad (3)$$

donde los pesos λ_k cumplen la condición:

$$\sum_{k=j-m}^{j+m} \lambda_k = 1. \quad (4)$$

$I^*(\omega_j)$ es el periodograma de Lomb-Scargle suavizado para la frecuencia ω_j .

Existen diferentes modos de elegir los pesos. Si todos tienen el mismo valor, sujeto a la condición de la ecuación (4), se tiene la media aritmética. Otros tipos de suavizado pueden verse en Chatfield (1991).

El número de frecuencias en las cuales se evalúa el periodograma de Lomb-Scargle es generalmente ma-

yor que en el caso de muestreo regular con intervalo de muestreo Δ . En este último caso, la frecuencia más alta para las cuales se puede evaluar la potencia espectral es f_n , la frecuencia de Nyquist (Papoulis, 1984), que se define como la frecuencia que corresponde a un período igual a dos veces el intervalo de muestreo:

$$f_n = \frac{1}{2\Delta}. \quad (5)$$

En el caso de muestreo irregular, hay varias frecuencias que se pueden tomar como referencia. Una es la frecuencia de Nyquist equivalente, que consiste en considerar que los N datos están regularmente espaciados, en cuyo caso se obtendría un intervalo de muestreo

$$\Delta_1 = \frac{t_N - t_1}{N - 1}, \quad (6)$$

y la frecuencia de Nyquist sería:

$$f_1 = \frac{1}{2\Delta_1}. \quad (7)$$

Otra frecuencia de interés es la frecuencia equivalente a la de Nyquist considerando la distancia mínima:

$$\Delta_m = \min(t_{i+1} - t_i); i = 1, \dots, N-1, \quad (8)$$

para la cual la frecuencia de Nyquist valdría:

$$f_m = \frac{1}{2\Delta_m}. \quad (9)$$

No tiene sentido investigar frecuencias mayores a esta frecuencia f_m .

Todas las consideraciones anteriores han de tenerse en cuenta al estimar el periodograma de Lomb-Scargle. Por ejemplo, el programa descrito en Pardo-Igúzquiza y Rodríguez-Tovar (2012), pide especificar la frecuencia máxima f_M a investigar y el número de frecuencias a calcular en el intervalo $(0, f_M)$.

Un aspecto muy importante a considerar es la evaluación de la significación estadística del periodograma de Lomb-Scargle estimado. Bajo la hipótesis nula de ruido blanco gaussiano, el periodograma sigue una distribución exponencial (Brockwell y Davis, 1991). Sin embargo, el periodograma de Lomb-Scargle se evalúa con un sobremuestreo en las frecuencias, lo que significa que el número de frecuencias para las cuales se estima la potencia espectral es mayor que con la transformada discreta de Fourier para el caso

de muestreo regular. Este sobremuestreo implica una pérdida en el número de grados de libertad (esto es, el número de valores de la potencia espectral estimada que pueden considerarse como independientes). Adicionalmente, el suavizado del periodograma estimado introduce una pérdida de grados de libertad adicional ya que hace que los valores del periodograma estimados para frecuencias vecinas estén más correlacionados. En este caso, tratar de obtener una fórmula analítica que exprese la significación estadística del espectro estimado puede ser bastante complicado. Resulta mucho más efectivo utilizar un método de Monte Carlo con cálculo intensivo por ordenador. Uno de estos métodos es el test de permutación (Efron y Tibshinari, 1994; Good, 2000; Pardo-Igúzquiza y Rodríguez-Tovar, 2000). En este método, el periodograma de Lomb-Scargle se calcula no solo para la serie original sino para un gran número de permutaciones aleatorias de la serie temporal original. Una serie permutada se define para las mismas posiciones originales $t_1 < t_2 < \dots < t_N$ y donde los mismos datos originales se asignan de modo aleatorio a las diferentes posiciones. De este modo, lo más probable es que la permutación destruirá cualquier señal o patrón cíclico que estuviera presente en la serie original y la hipótesis nula del test de permutación es que las permutaciones aleatorias son un ruido blanco con la misma distribución (histograma) que los datos experimentales, los cuales no necesitan seguir una distribución gaussiana.

El periodograma de Lomb-Scargle se calcula para cada una de las permutaciones, donde el número de permutaciones S es un valor alto (1000 como mínimo). De este modo el nivel de significación (ASL) se calcula como:

$$ASL(\omega_j) = \frac{1}{S} \sum_{i=1}^S J_i(\omega_j), \quad (10)$$

$$J_i(\omega_j) = \begin{cases} 1 & \text{si } I^*(\omega_j) \leq \tilde{I}_i^*(\omega_j) \\ 0 & \text{en caso contrario} \end{cases} \quad (11)$$

donde $I^*(\omega_j)$ es el periodograma suavizado de Lomb-Scargle de la secuencia original para la frecuencia ω_j , y $\tilde{I}_i^*(\omega_j)$ es el periodograma suavizado de la i -ésima permutación para la frecuencia ω_j . El nivel de confianza se define, en porcentaje, como:

$$ACL(\omega_j) = 100[1 - ASL(\omega_j)] \quad (12)$$

Estudio de casos de muestreo irregular

La Fig. 1 muestra una serie simulada compuesta de señal más ruido gaussiano de media cero y varianza 3. La serie consta de 500 datos con muestreo regular con interespaciado unidad $\Delta = 1$ año. La señal consta de tres componentes cíclicas (sinusoides) con las siguientes características: amplitudes $A_1 = 1$, $A_2 = 1$ y $A_3 = 1$, frecuencias $f_1 = 0.008$, $f_2 = 0.05$, $f_3 = 0.4$ y fases $\varphi_1 = 45^\circ$, $\varphi_2 = 22.5^\circ$ y $\varphi_3 = 15^\circ$. La frecuencia $f_1 = 0.008$ corresponde a una baja frecuencia con período 125 años que se repite 4 veces en la secuencia. La frecuencia $f_2 = 0.05$ corresponde a una frecuencia media con período 20 años que se repite 25 veces en la secuencia. La frecuencia $f_3 = 0.4$ corresponde a una alta frecuencia con período 2.5

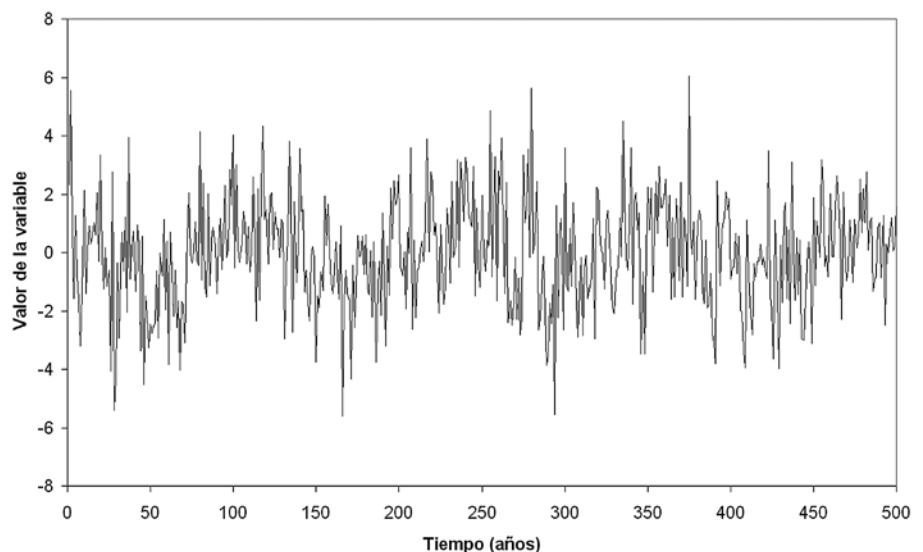


Figura 1. Serie simulada igual a la señal más ruido son una relación señal a ruido de 0.5.

Figure 1. Simulated time series with two components: signal and noise. The signal-to-noise ratio is 0.5.

años que se repite 200 veces en la secuencia. La varianza de la señal es de 1.5, aportando cada componente cíclica una varianza de 0.5, y la varianza del ruido vale 3 con lo que se obtiene una relación de señal a ruido de 0.5, esto es, la relación entre la varianza de la señal y la varianza del ruido. La varianza total de la serie es de 4.2 y por lo tanto cada componente cíclica viene a explicar un 12 % de la varianza total. El análisis espectral de la secuencia de 500 datos con intervalo de muestreo constante puede realizarse utilizando el periodograma de Lomb-Scargle y los resultados se muestran en la Fig. 2A. La Fig. 2B muestra el nivel de confianza de los picos estimados en la Fig. 2A. Puede observarse como los picos son perfectamente detectados en este caso, correspondiente a la serie con 500 datos y con intervalo de muestreo constante. Pero los casos que nos interesan son los de muestreo irregular, como los ejemplificados en la Fig. 3. Estos casos se obtienen a partir de la serie completa representada en la Fig. 1. La Fig. 3A presenta una serie de 300 datos con un muestreo regular pero con varios hiatos espaciados regularmente a lo largo de la secuencia y la Fig. 3B presenta una serie de 300 datos con muestreo a intervalo constante pero con un gran hiato central. Finalmente, la Fig. 3C presenta una serie de 200 datos con muestro aleatorio.

La Fig. 4A muestra el periodograma de Lomb-Scargle de la serie con hiatos de la Fig. 3A. Puede observarse como se han identificado los tres ciclos de la señal pero parte de su potencia espectral se ha derramado (leakage) a frecuencias vecinas. El derrame espectral implica que los picos de las señales no aparecen tan acentuados como deberían ya que parte de su valor se atribuye a frecuencias vecinas (derrame espectral), lo que puede provocar o un ensanchamiento del pico o la aparición de picos espurios vecinos al pico verdadero de la señal. Esto se produce porque la función experimental

puede verse como el producto de la serie original por una ventana de muestreo como la representada en la Fig. 5A. Es esta ventana de muestreo la que introduce el derrame espectral. La serie observada en la Fig. 3A puede considerarse como el producto de la serie original de la Fig. 1 por la ventana de la Fig. 5A y por consiguiente el periodograma estimado es la convolución de la transformada de Fourier de la señal original con la transformada de Fourier de la ventana que se muestra en la Fig. 5B. Como puede observarse en la Fig. 5B, aparece un lóbulo secundario para la frecuencia 0.01, por lo que se producirán importantes picos secundarios debido al derrame espectral en el periodograma de la serie de la Fig. 3A para frecuencias 0.008 ± 0.01 , 0.05 ± 0.01 y 0.4 ± 0.01 como puede verse claramente en la Fig. 4A. Aunque el nivel de confianza de estos ciclos espurios es muy alto como se aprecia en la Fig. 4B, no llega a ser tan alto como los de las frecuencias originales de la señal. Además como se ha indicado, su efecto es predecible y fácilmente interpretable en este caso de ciclos para frecuencias bien separadas.

La Fig. 6A muestra el periodograma de Lomb-Scargle de la serie con hiatos de la Fig. 3B. Puede observarse como se han identificado los tres ciclos de la señal pero aparecen con mayor anchura que en el caso anterior. Asimismo se reconoce también este efecto en el nivel de confianza de la Fig. 6B. Esto se produce por el efecto de la ventana de datos que en este caso es la mostrada en la Fig. 7A. El efecto frecuencial de esta ventana de datos puede interpretarse a partir de su transformada de Fourier representada en la Fig. 7B. Puede verse como el lóbulo secundario está a una frecuencia muy pequeña de 0.002 lo que implica que el derrame espectral no se manifiesta como picos secundarios sino como un incremento en la anchura de los picos para las frecuencias de interés.

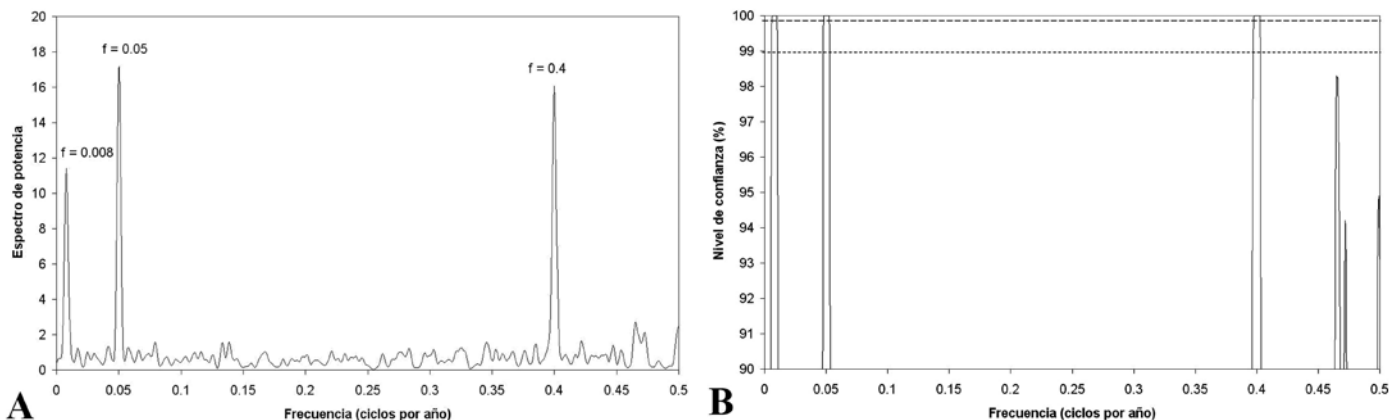


Figura 2. A: Periodograma de Lomb-Scargle de la serie mostrada en la Fig. 1. B: Nivel de confianza del espectro de la Fig. A obtenido mediante el test de permutación.

Figure 2. A: Lomb-Scargle periodogram of the series shown in Figure 1. B: Confidence level of the power spectrum in Figure A calculated using the permutation test.

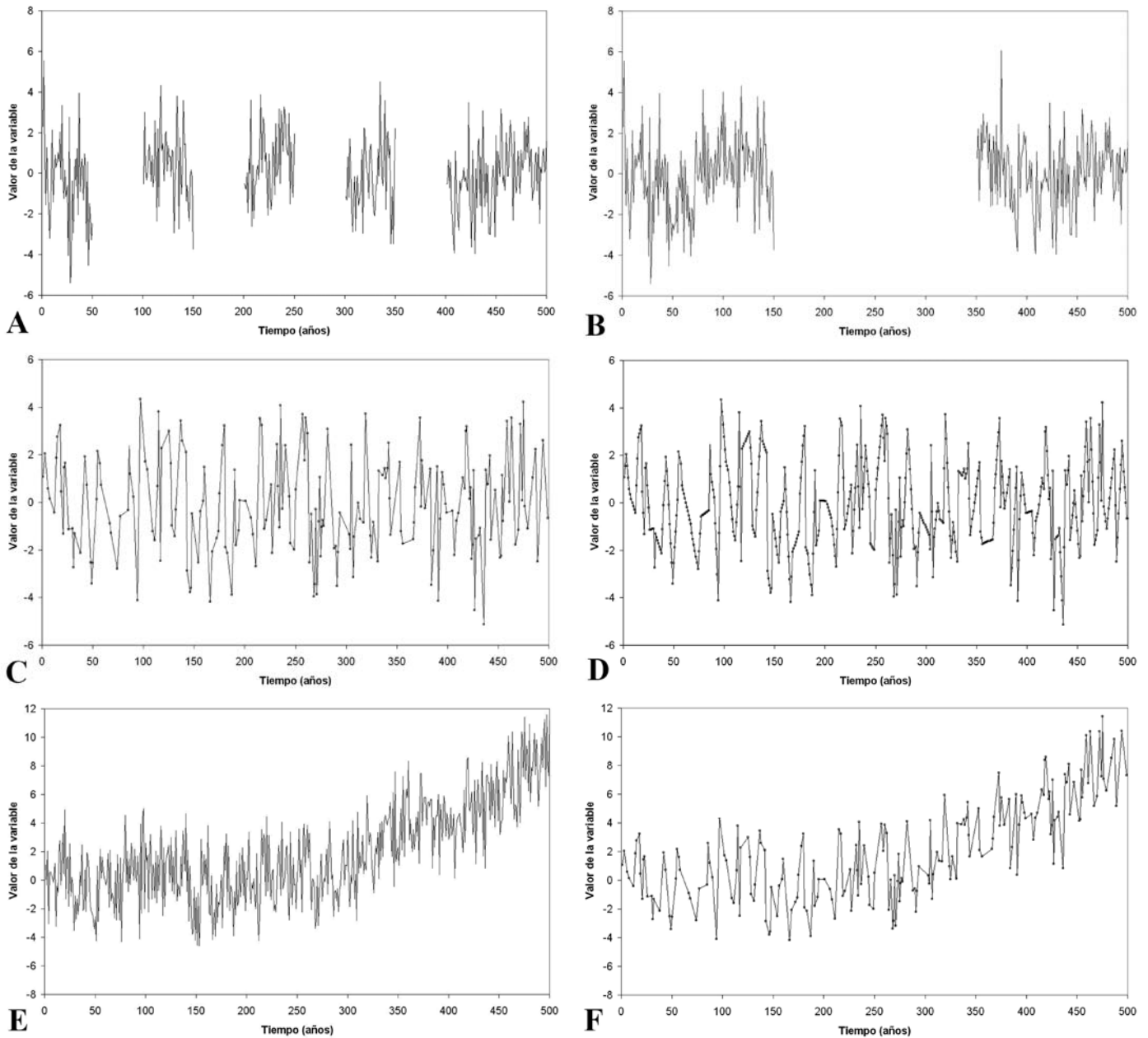


Figura 3. A: Serie temporal de 300 datos con muestreo con intervalo constante pero con hiatos a intervalos regulares (*gapped data*). B: Serie temporal de 300 datos con muestreo a intervalo constante pero con un hiato en la parte central. C: Serie temporal de 200 datos con muestreo aleatorio. D: Como la serie 3C pero interpolada linealmente para obtener una serie con muestreo regular a intervalo constante. E: Secuencia simulada similar a la de la Fig. 3A pero con una tendencia lineal en la segunda mitad de la secuencia. F: Secuencia de 200 datos con localización aleatoria tomados de la secuencia en 3E.

Figure 3. A: Time series of 300 data with a constant sampling interval but with hiatuses at constant intervals (*gapped data*). B: Time series of 300 data with constant sampling interval but with a hiatus in the middle. C: Time series of 200 data with random sampling. D: Time series in Figure 3C completed with linear interpolation to obtain even sampling with a constant sampling interval. E: Simulated sequence similar to the one in Figure 3A but with a linear trend in the second half of the sequence. F: Sequence of 200 data and random sampling from Figure 3E.

Hay que considerar que el hiato de la Fig. 3B tiene la misma anchura que la suma de hiatos más pequeños de la Fig. 3A. Sin embargo, salvo que existan ciclos con frecuencias muy próximas, el efecto del hiato único es menos perjudicial que el de los hiatos más pequeños equiespaciados.

La Fig. 8A muestra el periodograma de Lomb-Scargle de la serie con datos irregularmente distribuidos de la Fig. 3C y la Fig. 8B muestra el nivel de confianza obtenido. Este es el caso de muestreo aleatorio donde 200 datos son elegidos al azar de la secuencia original de 500 datos. Se observa como los picos son perfec-

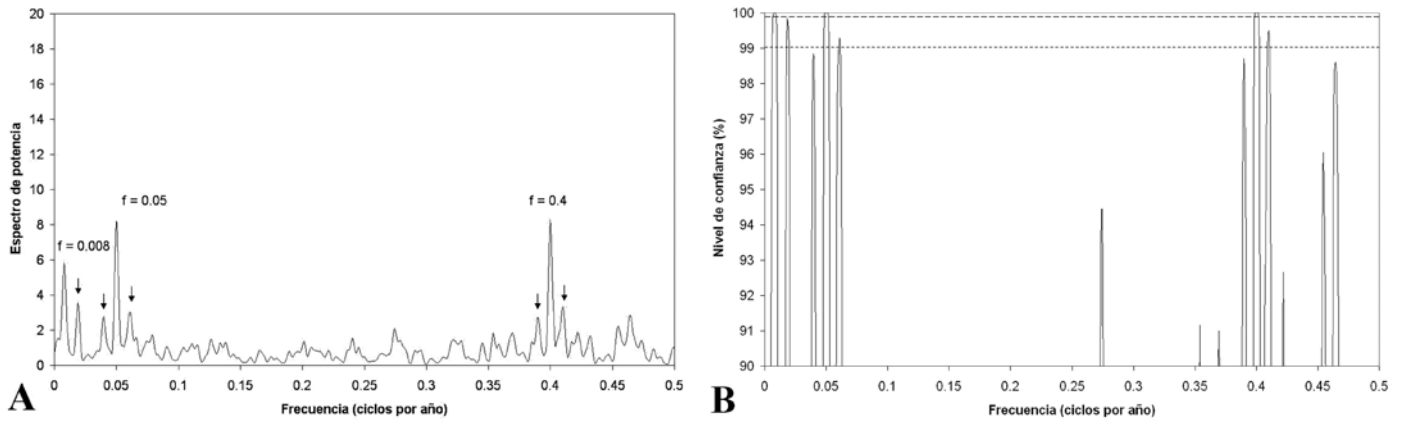


Figura 4. A: Periodograma Lomb-Scargle de la serie con hiatos mostrada en la Fig. 3A. Las flechas indican picos secundarios como consecuencia del muestreo irregular con hiatos. B: Nivel de confianza del periodograma mostrado en A.
Figure 4. A: Lomb-Scargle periodogram of the time series containing the hiatuses shown in Figure 3A. The arrows point to secondary peaks from this particular uneven sampling. B: Confidence level of the power spectrum shown in A.

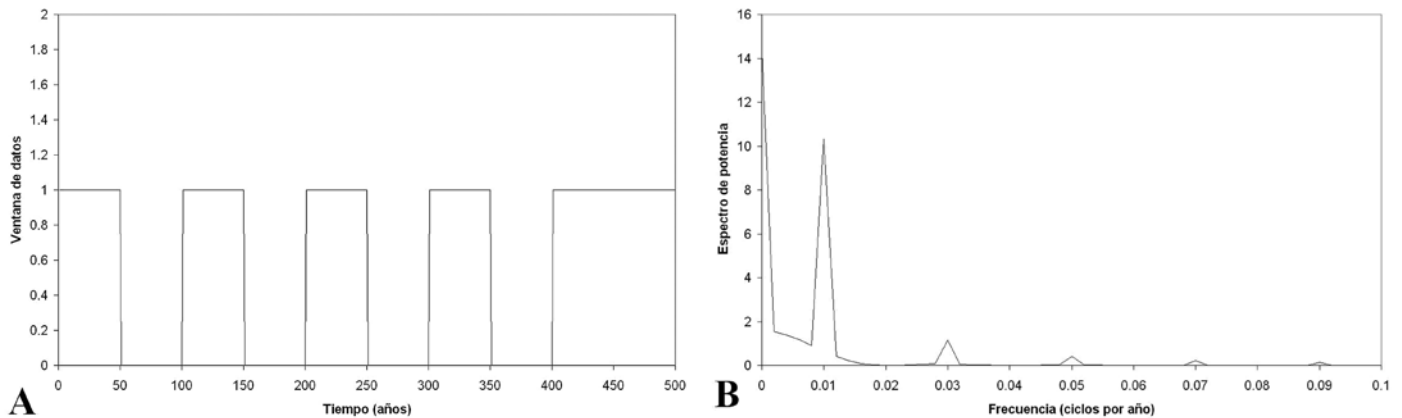


Figura 5. A: ventana de datos que aplicado a la secuencia original de la Fig. 1 produce la secuencia observada con hiatos de la Fig. 3A. B: transformada de Fourier de la ventana de datos que produce el derrame espectral para una frecuencia de 0.01 a ambos lados de cualquier ciclo presente en la serie a analizar.
Figure 5. A: Sampling window applied to the time series in Figure 1 to obtain the sequence with hiatuses shown in Figure 3A. B: Fourier transform of the sampling window responsible for the spectral leakage with a side lobe at a frequency of 0.001 on both sides of any spectral peak in the signal.

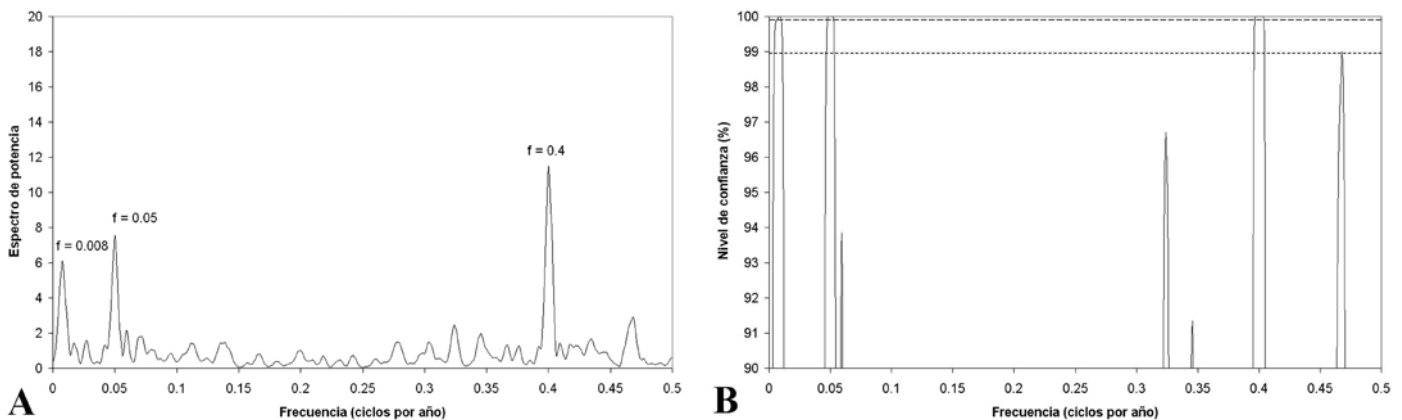


Figura 6. A: Periodograma Lomb-Scargle de la serie con hiatos mostrada en la Fig. 3B. B: Nivel de confianza del periodograma mostrado en A.
Figure 6. A: Lomb-Scargle periodogram of the time series with one hiatus shown in Figure 3B. B: Confidence level of the power spectrum shown in A.

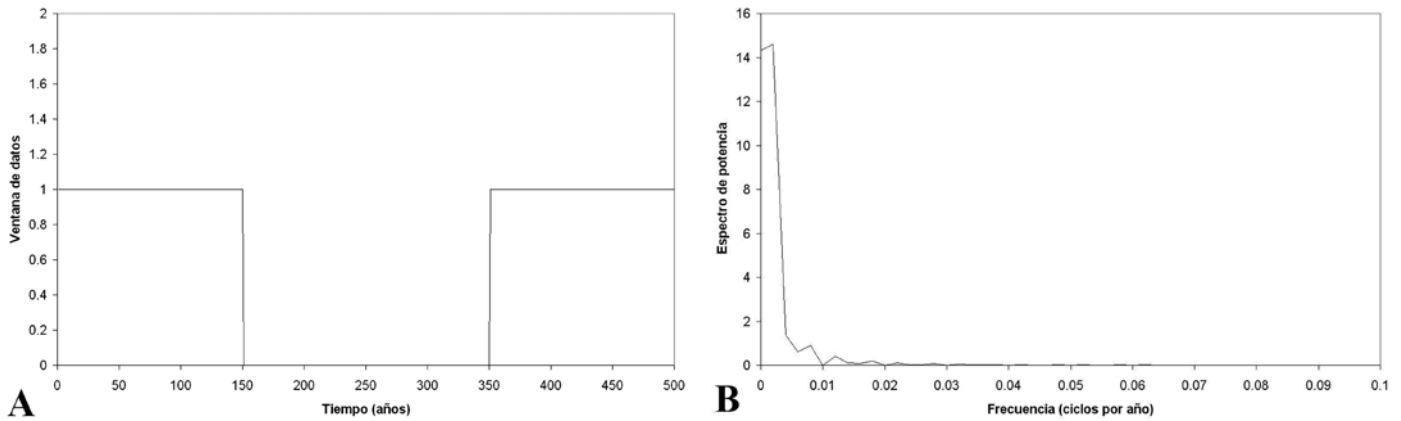


Figura 7. A: Ventana de datos que aplicado a la secuencia original de la Fig. 1 produce la secuencia observada con un hiato de la Fig. 3B. B: Transformada de Fourier de la ventana de datos que produce el derrame espectral para una frecuencia de 0.002 a ambos lados de cualquier ciclo presente en la serie a analizar lo que implica una mayor anchura de los picos detectados.

Figure 7. A: Sampling window applied to the time series in Figure 1 to obtain the sequence with hiatuses shown in Figure 3B. B: Fourier transform of the sampling window responsible for the spectral leakage with a side lobe at a frequency of 0.002 on both sides of any spectral peak in the signal.

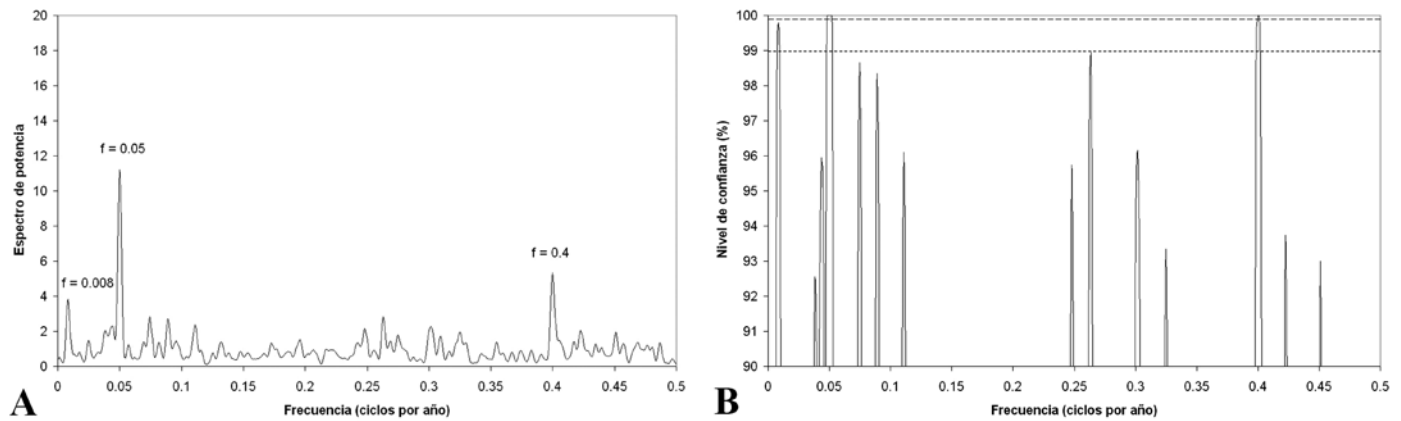


Figura 8. A: Periodograma Lomb-Scargle de la serie con datos irregulares mostrada en la Fig. 3D. B: Nivel de confianza del periodograma mostrado en A.

Figure 8. A: Lomb-Scargle periodogram of the time series with one hiatus shown in Figure 3D. B: Confidence level of the power spectrum shown in A.

tamente detectados y en mejores condiciones que en los casos con hiatos o huecos como los mostrados en las Fig. 3A y 3B.

Interpolación (caso N = 200)

En general, salvo el caso de una serie con muestreo uniforme donde algunos datos con localización aleatoria se han perdido (*missing at random*) y donde estos datos suponen menos del 10% de la serie completa, la interpolación o imputación de los valores faltantes es no es conveniente ya que modifica el contenido espectral de la serie original. El efecto de la interpolación se ha ilustrado con la serie de muestreo aleatorio con 200 datos. La Fig. 3D muestra la serie de la Fig. 3C interpolada linealmente para obtener una serie con intervalo de muestreo constante, mientras

las Fig. 9A y 9B muestran el periodograma de Lomb-Scargle y el nivel de confianza de los picos estimados. Puede apreciarse como el pico de las altas frecuencias ha desaparecido y, a pesar de que se identifican el pico de las bajas frecuencias y el de las frecuencias medias, aparecen ciclos espurios con un nivel de confianza similar a los verdaderos, lo que podría llevar a interpretaciones erróneas. Por el efecto de la interpolación, se ha perdido mucha información con respecto a utilizar el periodograma de Lomb-Scargle aplicada a la serie sin interpolación.

Presencia de una tendencia (caso N = 200)

En el apartado metodológico se ha indicado como una de las suposiciones para la aplicación del análisis espectral es que la serie es estacionaria de segun-

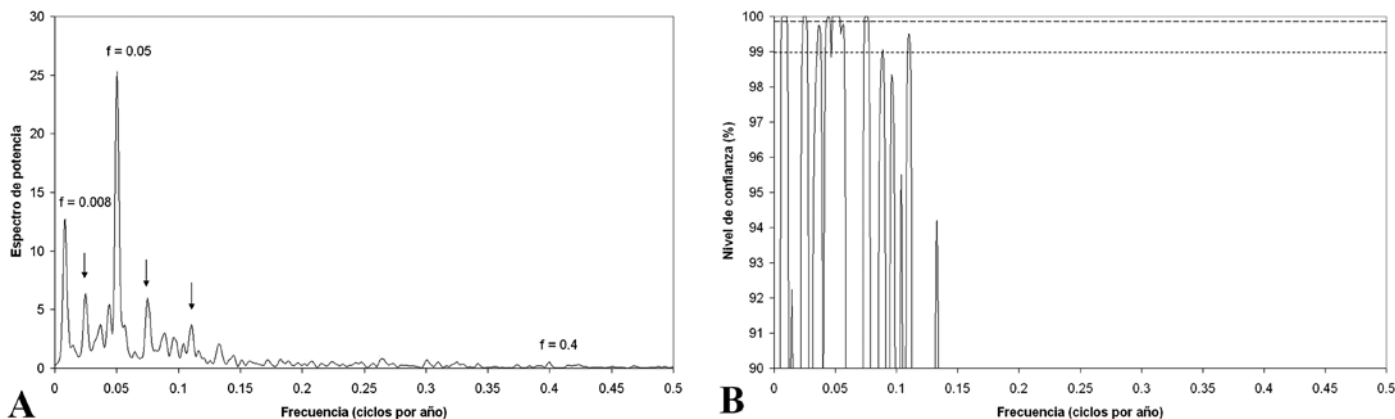


Figura 9. A: Periodograma Lomb-Scargle de la serie interpolada mostrada en Fig. 3D. B: Nivel de confianza del periodograma mostrado en A.

Figure 9. A: Lomb-Scargle periodogram of the time series with one hiatus shown in Figure 3D. B: Confidence level of the power spectrum shown in A.

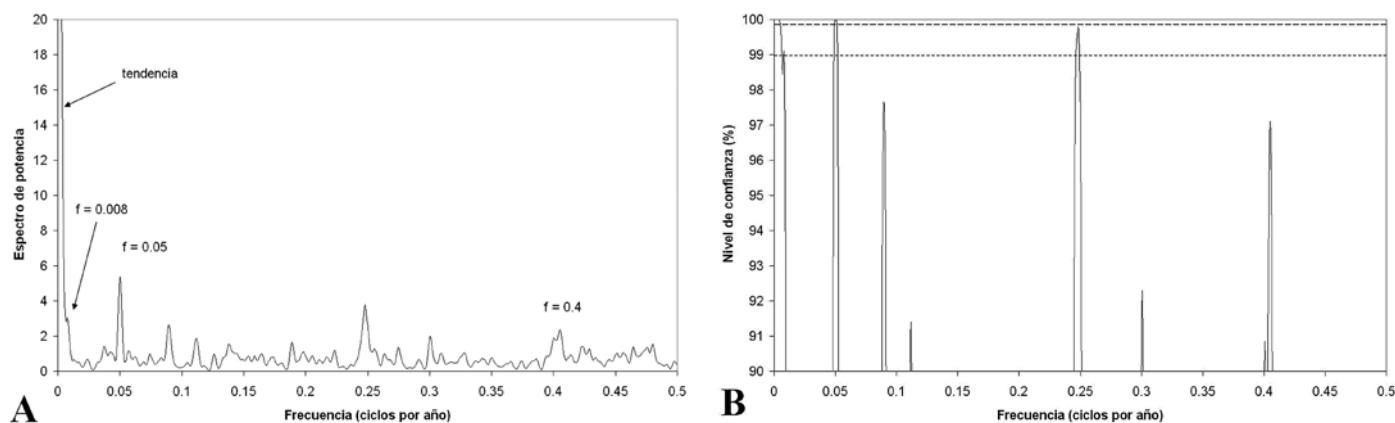


Figura 10. A: Periodograma Lomb-Scargle de la serie interpolada mostrada en Fig. 3F. B: Nivel de confianza del periodograma mostrado en A.

Figure 10. A: Lomb-Scargle periodogram of the time series with one hiatus shown in Figure 3F. B: Confidence level of the power spectrum shown in A.

do orden. Esta suposición no se cumple en particular cuando existe una tendencia en la media. Un ejemplo puede verse en la Fig. 3E con una tendencia lineal en la segunda mitad de la serie y con una serie irregular de 200 datos (Fig. 3F) que corresponden al mismo muestreo que en la Fig. 3C.

La Fig. 10A muestra el periodograma de Lomb-Scargle de la serie con datos irregularmente distribuidos de la Fig. 3F y la Fig. 10B muestra el nivel de confianza obtenido. El efecto más negativo en el espectro estimado es la introducción de una frecuencia espuria altamente significativa en la frecuencia 0.25. En este caso de tendencias lineales es posible efectuar una estimación de la tendencia utilizando un método no paramétrico como una media móvil. Esta tendencia estimada se ha mostrado en la Fig. 11A y la serie después de descontar la tendencia se ha mostrado en la Fig. 11B. Cuando se aplica de nuevo el periodograma

de Lomb-Scargle a la serie sin tendencia se obtiene el resultado mostrado en la Fig. 12A y con el nivel de confianza mostrado en la Fig. 12B. Se observa como en estos casos en los que se suprime exitosamente la tendencia, se obtienen los resultados correctos, identificándose los tres ciclos presentes en la señal.

Análisis espectral de un indicador paleoclimático en un espeleotema

Los espeleotemas, especialmente estalagmitas, están ganando relevancia para obtener indicadores paleoclimáticos en ámbitos continentales. Como ejemplo se ha tomado el espesor en milímetros de bandas anuales en una estalagmita de la cueva Carlsbad en Nuevo México (EEUU), de acuerdo a Polyak y Asmerom (2001). La serie tiene 1896 datos para un interva-

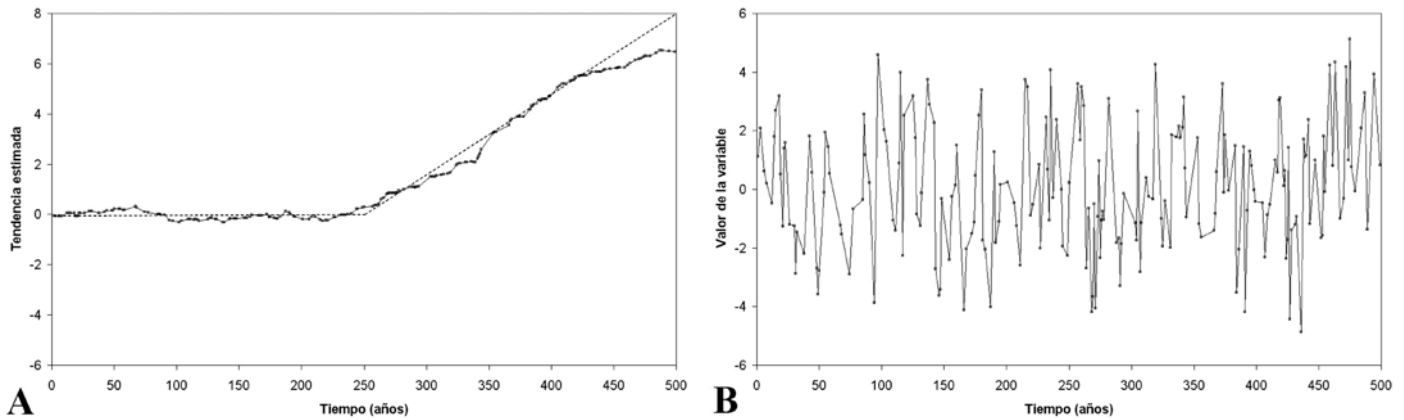


Figura 11. A: Tendencia estimada a partir de la secuencia mostrada en la Fig. 3F. B: Secuencia sin tendencia al descontar la tendencia mostrada en A de la secuencia con tendencia mostrada en la Fig. 3F.

Figure 11. A: Estimated trend of the time series shown in Figure 3F. B: Detrended sequence obtained by removing the trend in A from the time series in Figure 3F.

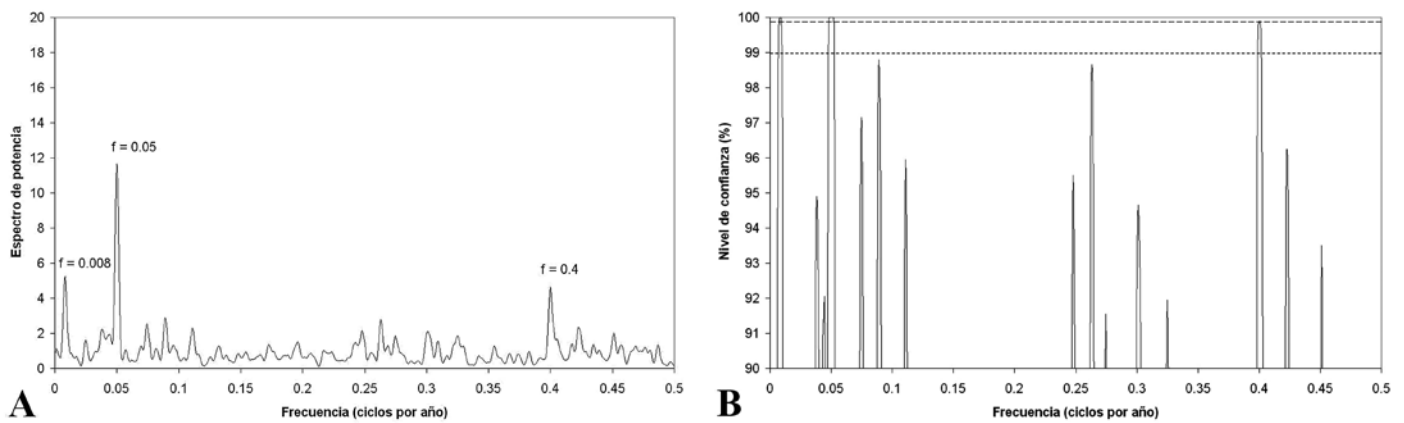


Figura 12. A: Periodograma Lomb-Scargle de la serie sin tendencia mostrada en Fig. 11B. B: Nivel de confianza del periodograma mostrado en A.

Figure 12. A: Lomb-Scargle periodogram of the detrended time series with one hiatus shown in Figure 11B. B: Confidence level of the power spectrum shown in A.

lo de 2671 años, desde el 161 al 2832 años antes del presente (Fig. 13). Puede observarse un hiato claro en torno a los 600 años antes del presente, junto con otros más pequeños, totalizando la ausencia de 775 bandas anuales. La Fig. 14A representa la ventana de datos que, aplicada a la serie completa, proporciona la serie observada con hiatos en la Fig. 13 y su transformada de Fourier se ha representado en la Fig. 14B. Es de destacar como los lóbulos laterales no son muy importantes con respecto al lóbulo central, por lo que no es de esperar un derrame espectral significativo que complique la interpretación del espectro de la serie original. Utilizando una media móvil de 200 años se obtiene la tendencia estimada que se muestra en la Fig. 15A y que se sustraerá de la secuencia original de la Fig. 13 para obtener la muestra sin tendencia de la Fig. 15B sobre la que se ha calculado el periodograma de Lomb-Scargle. El

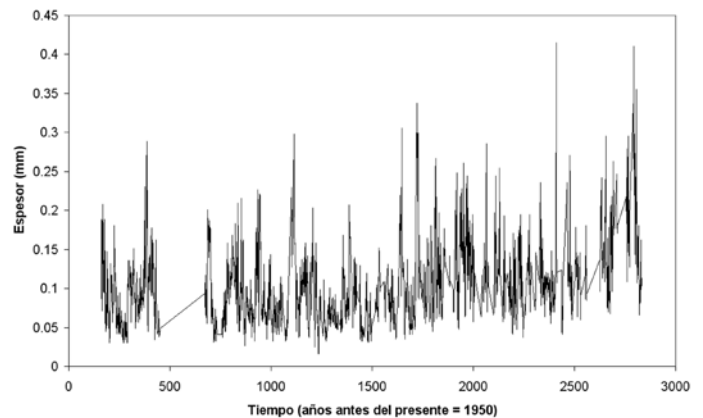


Figura 13. Espesor de bandeo anual en una estalagmita de la cueva Carlsbad (Nuevo México, EEUU) para los últimos 3000 años antes del presente. Nótese los hiatos existentes especialmente el situado entorno a hace unos 600 años antes del presente.

Figure 13. Annual band thickness of a stalagmite in Carlsbad Cave (New Mexico, USA) for the last 3000 years BP. Uneven sampling because of hiatuses, the main one around 600 years BP.

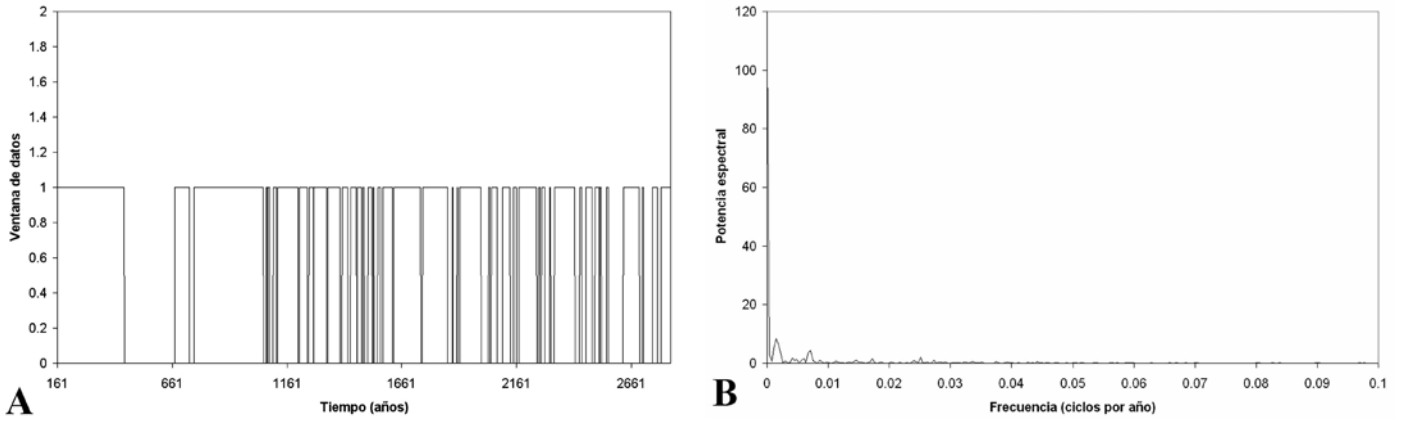


Figura 14. A: Ventana de datos que produce la secuencia observada con hiatos de la Fig. 13. B: Transformada de Fourier de la ventana de datos que produce el derrame espectral para diversas frecuencias.
Figure 14. A: Sampling window applied to the complete time series to obtain the sequence with hiatuses shown in Figure 13. B: Fourier transform of the sampling window responsible for weak spectral leakage.

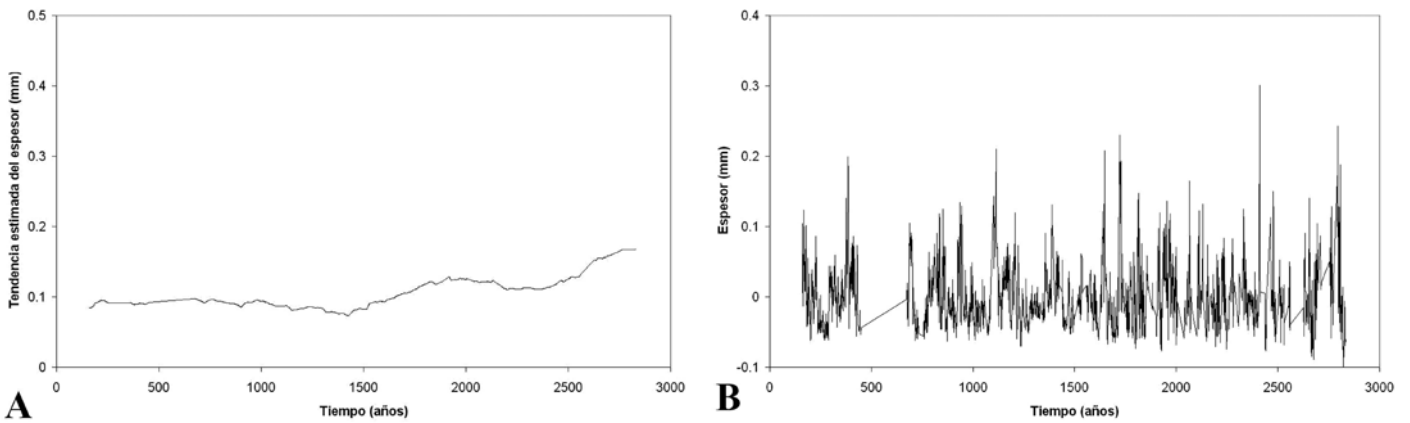


Figura 15. A: Tendencia estimada por una media móvil de anchura 200 años. B: Serie de espesores después de sustraer la tendencia.
Figure 15. A: Estimated trend using a running mean of 200 years. B: Experimental time series of thicknesses after detrending.

espectro de potencia estimado se muestra en la Fig. 16A y su nivel de confianza en la Fig. 16B. Se consideran como significativos los picos cuyas frecuen-

cias corresponden a períodos de 200, 80, 48, 26.6, 22, 16 y 11 años. Una interpretación preliminar de los ciclos podría ser la siguiente. El ciclo de 11 años se

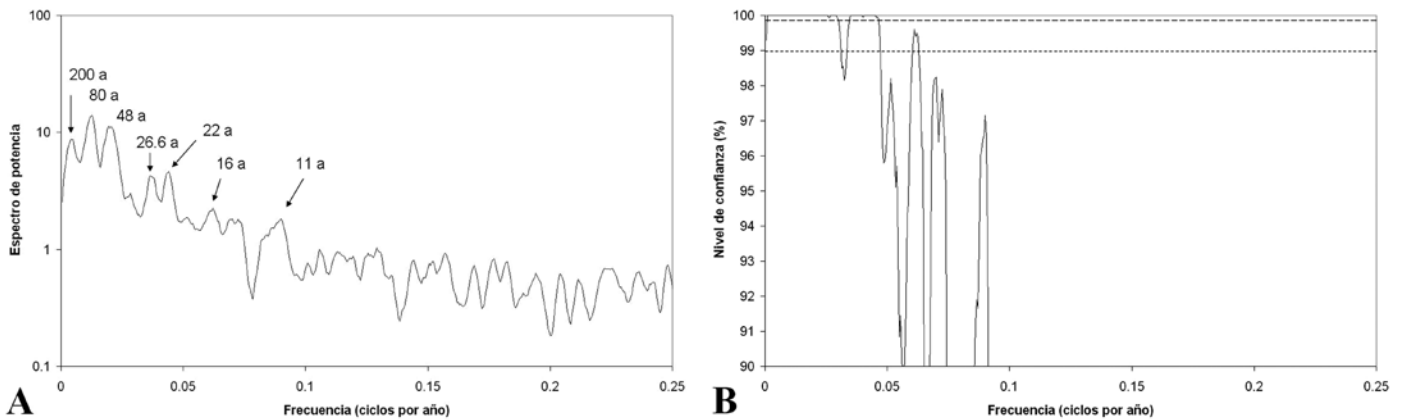


Figura 16. A: Periodograma Lomb-Scargle de la serie mostrada en Fig. 15B. B: Nivel de confianza del periodograma mostrado en A.
Figure 16. A: Lomb-Scargle periodogram of the time series with one hiatus shown in Figure 15B. B: Confidence level of the power spectrum shown in A.

puede correlacionar con el ciclo del número de manchas solares (ciclo de Schwabe o ciclo Solar) y el ciclo de 22 años con el ciclo magnético de las manchas solares (ciclo de Hale). Picos con períodos de 26.9 y 78.7, muy próximos a los de 26.6 y 80 años respectivamente, se han detectado por Pardo-Igúzquiza *et al.* (1994) en el análisis de los datos de varvas lacustres del Eoceno obtenidos de Davis (1973). El ciclo de 16 años se puede relacionar con el pico de 16-17 años que Mann y Lees (1996) asocian a variaciones del nivel del lago Great Salt. El ciclo de 48 años también se ha encontrado en otras estalagmitas en Europa muy distantes de Nuevo México. El ciclo de 200 años podría corresponder a un ciclo de la actividad solar de más amplio rango que los anteriores. Este caso de estudio ha ilustrado la aplicación del periodograma de Lomb-Scargle para una serie real de un espeleotema con hiatos y con una tendencia en la media. Los ciclos estimados como estadísticamente significativos se han podido correlacionar con ciclos en la variabilidad solar.

Conclusiones

Muchas de las secuencias utilizadas en el análisis cicloestratigráfico y procedentes del registro geológico se obtienen a partir de un muestreo irregular. En este tipo de muestreo el espaciado entre datos consecutivos no es constante. En estos casos la interpolación espacial no es una buena estrategia de trabajo ya que altera el contenido espectral de la secuencia. Una estrategia más conveniente es el utilizar un método de estimación espectral como el periodograma de Lomb-Scargle, que analiza directamente los datos sin intervalo de muestreo constante. La conveniencia y fiabilidad de su utilización se demuestra en diferentes casuísticas, aplicando distintos tipos de muestreos irregulares y su efecto sobre el análisis espectral, así como evaluando los efectos de una tendencia. Se ha mostrado cómo el análisis de la transformada de Fourier de la ventana de muestreo, conjuntamente con la significación estadística de los picos espectrales evaluada por el test de permutación, pueden utilizarse para interpretar el espectro estimado y su relación con las posibles causas físicas que los pudieron originar. La aplicación de esta metodología a una serie real obtenida a partir de los espesores de las bandas anuales en un espeleotema, ha permitido caracterizar los picos espectrales de mayor significación, e interpretar de manera preliminar su génesis en relación con ciclos de variabilidad solar.

Agradecimientos

Este trabajo ha sido financiado por los proyectos de investigación CGL2010-15498 y CGL2012-33281 del Ministerio de Economía y Competitividad y CGL2008-03007 del Ministerio de Ciencia e Innovación así como P08-RNM-03715 de la Junta de Andalucía. Los autores agradecen las sugerencias de los revisores (Juan José Egozcue y Ricardo Olea) que han servido para mejorar la versión final de este trabajo.

Referencias

- Babu, P. and Stoika, P. 2010. Spectral analysis of nonuniformly sampled data – a review. *Digital Signal Processing*, 20, 359-378.
- Brockwell, P.J. and Davis, R.A. 1991. *Times series: theory and methods*. Second Edition. Springer Verlag, New York, 577 pp.
- Chatfield, C. 1991. *The analysis of time series: An introduction*. Fourth edition. Chapman and Hall, London, 241 pp.
- Davis, J.C. 1973. *Statistics and data analysis in geology*. John Wiley & Sons, New York, 550 pp.
- Efron, B. and Tibshirani, R.J. 1994. *An introduction to the bootstrap*. Chapman and Hall, New York, 436 pp.
- Good, P. 2000. *Permutation test: A practical guide to resampling methods for testing hypothesis*. Second Edition. Springer, New York, 270 pp.
- Hasselmann, L. 1976. Stochastic climate models: Part I. Theory. *Tellus*, 28 (6), 473-485.
- 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.
- Lomb, N.R. 1976. Least squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39, 447-462.
- Mann, M.E. and Lees, J. 1996. Robust Estimation of Background Noise and Signal Detection in Climatic Time Series. *Climatic Change*, 33, 409-445.
- 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*. Second Edition. McGraw-Hill International 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. 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. 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.
- Park, J., Lindberg, C.R. and Vernon III, F.L. 1987. Multitaper spectral analysis of high frequency seismograms, *Journal Geophysical Research*, 92, 12675-12684.
- 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.
- 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.
- Stoica, P. and Sandgren, N. 2006. Spectral análisis of irregularly-sampled data: Paralleling the regularly-sampled data approaches. *Digital Signal Processing*, 16, 712-734.
- Weedon, G.P. 2003. *Times-series analysis and cyclostratigraphy: examining stratigraphic records of environmental cycles*. Cambridge University Press, Cambridge, 259 pp.

Recibido: diciembre 2012

Revisado: enero 2013

Aceptado: febrero 2013

Publicado: abril 2013

