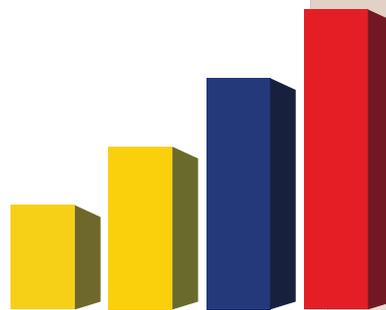


Analítica

Estimación espectral de datos ambientales no equiespaciados vía el periodograma suavizado de Lomb-Scarle. Una breve revisión.

Spectral estimation of unevenly environmental data via the smoothed Lomb-Scargle periodogram. A short review.

Josué M. Polanco-Martínez



www.ecuadorencifras.gob.ec

ESTIMACIÓN ESPECTRAL DE DATOS AMBIENTALES NO EQUIESPACIADOS VÍA EL PERIODOGRAMA SUAVIZADO DE LOMB-SCARGLE. UNA BREVE REVISIÓN

SPECTRAL ESTIMATION OF UNEVENLY ENVIRONMENTAL DATA VIA THE SMOOTHED LOMB-SCARGLE PERIODOGRAM. A SHORT REVIEW.

Josué M. Polanco Martínez

Basque Centre for Climate Change, Bilbao, España

josue.polanco@bc3research.org

Recibido: 25 de agosto 2014

Aceptado: 17 de octubre 2014

Resumen

El periodograma suavizado de Lomb-Scargle es una técnica de análisis espectral que se aplica de modo directo a series temporales no equiespaciadas. Aunque derivado originalmente para operar con series temporales astronómicas no equiespaciadas temporalmente [2, 17, 38, 39], a finales de los noventa fue adaptado por Schulz y Stattegger [43] en combinación con la técnica WOSA (*Welch-Overlapped-Segment-Averaging*) [49] para operar con series temporales ambientales (principalmente climáticas) no equiespaciadas temporalmente. Un poco más tarde, Schulz y Mudelsee [42] hicieron mejoras al trabajo de Schulz y Stattegger para tener en cuenta el tipo de ruido de fondo ("rojo") que suelen presentar las series ambientales. Debido a la necesidad de estimar el espectro suavizado a series temporales ambientales no equiespaciadas temporalmente, es necesario contar con información suficiente y de libre acceso sobre esta temática. Hoy por hoy, es posible encontrar una buena cantidad de publicaciones en inglés sobre este método (*v. gr.*, [23, 24, 28, 29, 42, 43]), pero hay una carencia de información en idioma español (salvo algunas excepciones, como Polanco-Martínez [31] y Pardo-Igúzquiza y Rodríguez-Tovar [27]). Por estas razones, en este artículo de revisión, se presenta de manera concisa toda la información pertinente para estimar el espectro suavizado de series temporales ambientales no equiespaciadas, mediante el periodograma de Lomb-Scargle y teniendo en cuenta el ruido de fondo rojo de las series ambientales.

Palabras clave: Análisis espectral, series temporales no equiespaciadas temporalmente, periodograma de Lomb - Scargle, transformada de Lomb-Scargle Fourier, frecuencia de Nyquist, ruido rojo.

Abstract

The smoothed Lomb-Scargle periodogram is a spectral analysis technique to be used directly to environmental unevenly spaced time series. The Lomb-Scargle periodogram was originally derived to be used with astronomical unevenly spaced time series [2, 17, 38, 39]. However, in the late nineties, Schulz and Stattegger [43] adapted this technique in combination with WOSA (*Welch-Overlapped-Segment-Averaging*) [49] to analyse environmental (mainly climatic) unevenly spaced time series. Later, Schulz and Mudelsee [42] brought some improvements in the work of Stattegger and Schulz [43] to take into account the kind of noise background ("red") of the environmental time series. Due to the need for estimating the smoothed spectrum from unevenly spaced environmental time series, it is necessary to have sufficient information about this promising statistical method on free access. Currently, there is an enough number of publications in English of this method (*v. gr.*, [23, 24, 28, 29, 42, 43]), but only few publications in Spanish (some exceptions are Polanco-Martínez [31] y Pardo-Igúzquiza y Rodríguez-Tovar [27]). Due to the potential use of this method, more information in Spanish is required. For these reasons, we are highly motivated to contribute with this review paper, which contains the state of the art to estimate the smoothed spectrum via the Lomb-Scargle periodogram from environmental unevenly spaced time series and taking into account a red noise background.

Keywords: Spectral analysis, unevenly spaced time series, Lomb-Scargle periodogram, Lomb-Scargle Fourier transform, Nyquist frequency, red noise.

Código JEL: C01, C14, C23.

1 Introducción

Hoy por hoy, el uso de la estadística en la investigación científica es una herramienta imprescindible en todas las áreas de la ciencia e ingeniería, y las ciencias ambientales (e.g., meteorología, climatología, hidrología, etc.), no son una excepción. De hecho, una parte considerable del tiempo que se dedica a investigación se invierte en cómo enfrentar las particularidades (como la pérdida de datos, la irregularidad temporal en el muestreo, la presencia de ruido, valores extremos, estacionalidad, no-estacionariedad, etc.) que presentan los diferentes tipos de datos ambientales (como los observacionales, los provenientes de medidas instrumentales, de reconstrucciones paleoambientales, de salidas de simulaciones numéricas, etc.), así como en averiguar el tipo de técnicas estadísticas más adecuadas para obtener información cuantitativa que nos permita ganar conocimiento sobre un determinado fenómeno ambiental que vayamos a estudiar [1, 23, 31, 47].

Entre las técnicas matemáticas más utilizadas para el análisis de series temporales ambientales, destacan las técnicas de análisis espectral mediante la transformada de Fourier. Esta herramienta es sumamente útil para averiguar la posible existencia de eventos periódicos o para indagar el comportamiento de una serie que vayamos a estudiar en el dominio de la frecuencia [9, 33, 51]. Existe una gran variedad de métodos de análisis espectral, pero la mayoría requiere que la serie esté equiespaciada temporalmente [9, 48, 51], y este es un requisito que no siempre se cumple con las series temporales ambientales. La forma clásica de enfrentar el hecho de que la serie no es equiespaciada es interpolar la serie original y convertirla en una serie temporalmente equiespaciada, para que posteriormente se estime su espectro mediante algunas de las técnicas estándar de análisis espectral. Sin embargo, la interpolación implica un conocimiento previo del comportamiento de la variable que se estudia, y suaviza los datos de tal modo que al calcular el espectro puede suprimir información espectral en las altas frecuencias; en razón de que altera la estimación del espectro, debería evitarse la interpolación [25, 42, 43, 48]. Otra solución consiste en utilizar métodos “no estándar” de análisis espectral que puedan aplicarse directamente a las series temporales ambientales, tales como el periodograma de Lomb-Scargle [17, 28, 29, 38, 39, 42, 43], el método CLEAN [14, 35], CLEANEST [5] o más recientemente, los métodos de Mathias *et al.* [21] o Nava [25].

El periodograma de Lomb-Scargle (PLS), también conocido como *Lomb-Scargle Fourier Transform* (LSFT), es una técnica de análisis espectral que puede ser aplicada a series temporales no equiespaciadas en el tiempo; forma parte de los métodos de análisis espectral por mínimos cuadrados (*Least-Squares Spectral Analysis* (LSSA) [17, 38, 39, 46]. El origen del PLS, por lo general, está asociado a Lomb [17] y Scargle [38]. Sin embargo, retornando un poco hacia el

pasado, en 1969, Barning [2] desarrolló una técnica similar y, poco más tarde, en 1971, Vanicek [46] propuso la idea de mínimos cuadrados para determinar el espectro.

Lomb [17] realizó una ampliación del trabajo de Barning [2], de tal modo que estudió las propiedades estadísticas del análisis de frecuencias por mínimos cuadrados para series no equiespaciadas en el tiempo. Entre los resultados estadísticos más importantes del trabajo de Lomb, están la demostración de que las alturas de los picos del espectro de un ruido Gaussiano tienen una distribución del tipo χ^2_2 y la noción de que existe una correlación entre la altura de este tipo de espectro para dos frecuencias cualesquiera f_1 y f_2 [17]. Un poco más recientemente, Scargle [38] demostró que el espectro obtenido del ajuste de una función armónica por mínimos cuadrados es equivalente al espectro obtenido vía el periodograma. Scargle también realizó un estudio muy detallado para estimar la significación estadística de la detección de una señal.

El Periodograma de Lomb-Scargle se hizo ampliamente conocido debido al trabajo de Press *et al.* [32], en parte debido a que publicó una codificación computacional de esta técnica. Originalmente, el PLS fue derivado para trabajar con series temporales astronómicas [2, 17, 38, 39]; pero a finales de los noventa fue adaptado por Schulz y Stattegger [43] en combinación con la técnica WOSA (*Welch-Overlapped-Segment-Averaging*) [49], para trabajar con series temporales ambientales (principalmente climáticas). Schulz y Stattegger [43] proporcionaron los niveles de significación estadística, tanto para el caso uni-variado o auto-espectro (utilizaron el intervalo de confianza de los métodos clásicos de análisis espectral [3, 43]) como para el bi-variado o coherencia espectral¹ (utilizaron el algoritmo desarrollado por Scannell y Carter [37] para establecer los niveles de significación estadística). Schulz y Stattegger [43] también proporcionaron un paquete computacional de nombre SPECTRUM para estimar el auto-espectro, el espectro cruzado, la coherencia, la fase de la coherencia así como análisis armónico. SPECTRUM es de uso gratuito y está disponible en <http://www.geo.uni-bremen.de/geomod/staff/mschulz/#software1>.

Un poco más tarde, Schulz y Mudelsee [42] hicieron mejoras al trabajo de Schulz y Stattegger [43], para tener en cuenta el tipo de ruido de fondo (“rojo”) que se presenta en las series ambientales en la determinación del auto-espectro. Ellos también liberaron en internet un paquete computacional de uso gratuito, de nombre REDFIT, disponible en <http://www.geo.uni-bremen.de/geomod/staff/mschulz/#software2>; éste tiene en cuenta dichas mejoras; es importante mencionar que REDFIT sólo aplica al caso uni-variante, esto es, solo estima el auto-espectro suavizado vía el periodograma de Lomb-Scargle.

Recientemente, Mudelsee *et al.* [24] presentaron algunas

¹Schulz y Stattegger no estimaron los intervalos de confianza para el espectro cruzado o “cross-spectrum”, esto debido a las dificultades que implica calcularlos desde el punto de vista estadístico y debido a que la utilización de la espectro cruzado no es de uso práctico en el análisis de series temporales [43].

adaptaciones a REDFIT, introduciendo el error de las dataciones de muestras paleoclimáticas en la estimación del espectro. Sin embargo, estas mejoras sólo aplican a series temporales ambientales cuyos tiempos de muestreo contienen errores considerables que podrían afectar la estimación del espectro. Un año después, Mudelsee [23] (capítulo 5) proporcionó, entre otras temáticas relacionadas con el análisis de series temporales ambientales, una amplia revisión de los principales conceptos teóricos y diversos algoritmos para la estimación espectral de series temporales ambientales no equiespaciadas vía el periodograma de Lomb-Scargle.

Por otro lado, un poco más recientemente, Pardo-Iguzquiza y Rodríguez-Tovar [28, 29] introdujeron una nueva forma de estimar la significación estadística de los picos espectrales obtenidos mediante el periodograma suavizado de Lomb-Scargle. Esta novel implementación consiste en el uso de un test de permutación para evaluar la significación estadística de los picos estimados mediante el PLS, tanto para el caso uni-variante como para el bi-variante [27, 28, 29]. Pardo-Iguzquiza y Rodríguez-Tovar [29] también liberaron un par de programas (SLOMBS y CSLOMBS) de dominio público para llevar a cabo estas tareas. Sin embargo, a pesar de la importancia, tanto desde el punto de vista metodológico como computacional, es importante tener en consideración que un test de permutación implica un ruido de fondo “blanco” (plano). Sin embargo, la amplitud espectral, en la mayoría de las series ambientales, disminuye exponencialmente a medida que decrecen sus frecuencias; esto implica más bien un ruido de fondo de tipo “rojo” [10, 13, 20, 42].

Debido a la necesidad de estimar el espectro en diferentes tipos de series temporales ambientales no equiespaciadas temporalmente –que provienen de diversas áreas de la ciencia y la ingeniería– es necesario contar con información oportuna sobre esta temática. En lo que respecta a información en lengua inglesa exista una gran cantidad de publicaciones al respecto (véase, por ejemplo, [23, 24, 28, 29, 42, 43]). Sin embargo, no sucede lo mismo en idioma español, con excepciones como las de Pardo-Iguzquiza y Rodríguez-Tovar [27] y Polanco-Martínez [31]. La carencia de esta información en lenguaje español ha sido una de las principales motivaciones para la escritura de este artículo.

Este artículo de revisión tiene como principal objetivo proporcionar los principales aspectos matemáticos y computacionales en la estimación espectral (auto-espectro) vía el periodograma suavizado de Lomb-Scargle, pero teniendo en consideración el tipo de ruido de fondo de las series temporales ambientales. El artículo se estructura de la siguiente manera: en la sección 1, se proporciona una breve introducción del análisis espectral, se justifica el porque la interpolación debería evitarse y se presenta el “estado del arte” de esta metodología. En la sección 2 se presentan las principales bases matemáticas del periodograma de Lomb-Scargle. La sección 3 discute sobre las diferencias entre el

periodograma crudo versus el suavizado. En la sección 4 se introduce el test de significación estadística utilizado cuando se estima el espectro mediante el PLS. La sección 5 presenta un par de ejemplos de la estimación del espectro vía el PLS haciendo uso del paquete computacional REDFIT. Por último, en la sección 6, se proporcionan las conclusiones. Adicionalmente, en este trabajo se presenta un par de anexos; el anexo A contiene los detalles técnicos para el uso del programa computacional REDFIT [41]; el anexo B, finalmente, presenta un programa en lenguaje R [34] para graficar las salidas de REDFIT.

2 Consideraciones matemáticas del periodograma de Lomb-Scargle

En esta sección, se presentan los principales aspectos teóricos del periodograma de Lomb-Scargle: la definición, las principales diferencias entre el periodograma clásico y el de Lomb-Scargle, la definición de frecuencia de Nyquist promedio, así como las frecuencias donde se estima el periodograma de Lomb-Scargle. Para mantener una nomenclatura adecuada y consistente, en este artículo se sigue principalmente la nomenclatura de Horne y Baliunas [15], Lomb [17], Press *et. al.* [32], Scargle [38] y Van-Donget *et. al.* [45].

2.1 Definición del periodograma de Lomb-Scargle

Para una serie temporal $X(t_n)$, donde $n = 1, 2, \dots, N$ (número de elementos de la serie temporal) con media cero y varianza σ^2 (*i. e.*, estacionaria de segundo orden) y sin la presencia de tendencia y de valores extremos, el periodograma de Lomb-Scargle (normalización de Lomb) se define [15, 17, 32] como:

$$P(\omega_i = 2\pi f_i) = \frac{\hat{S}_\tau(\omega_i)}{2\sigma^2} = \frac{1}{2\sigma^2} \left[\frac{XC_\tau^2(\omega_i)}{CC_\tau(\omega_i)} + \frac{XS_\tau^2(\omega_i)}{SS_\tau(\omega_i)} \right], \quad (1)$$

donde ω_i es la frecuencia angular y $f_i > 0$, $i = 1, 2, \dots, k$, forman el conjunto de frecuencias en las que se desea calcular el periodograma. Cada término de la relación (1) está definido por

$$XC_\tau(\omega_i) = \sum_{n=1}^N X(t_n) \cos[\omega_i(t_n - \tau(\omega_i))] \quad (2)$$

$$CC_\tau(\omega_i) = \sum_{n=1}^N \cos^2[\omega_i(t_n - \tau(\omega_i))] \quad (3)$$

$$XS_\tau(\omega_i) = \sum_{n=1}^N X(t_n) \sin[\omega_i(t_n - \tau(\omega_i))] \quad (4)$$

$$SS_\tau(\omega_i) = \sum_{n=1}^N \sin^2[\omega_i(t_n - \tau(\omega_i))] \quad (5)$$

donde τ , la constante que asegura la invariancia en el tiempo [38], está definida por

$$\tau(\omega_i) = \frac{1}{2\omega_i} \arctan \left[\frac{\sum_{n=1}^N \text{sen}(2\omega_i t_n)}{\sum_{n=1}^N \text{cos}(2\omega_i t_n)} \right]. \quad (6)$$

2.2 El periodograma clásico vs. el periodograma de Lomb-Scargle

La definición de periodograma clásico [38, 44] para el caso de muestreos equiespaciados en el tiempo, es :

$$p(f_i) = \frac{1}{N} \left(\left[\sum_{n=1}^N X(t_n) \cos(\omega_i t_n) \right]^2 + \left[\sum_{n=1}^N X(t_n) \sin(\omega_i t_n) \right]^2 \right), \quad (7)$$

donde $\omega_i = 2\pi f_i$, para $i=1, 2, \dots, k$. Los tiempos de muestreo están dados por $t_n = n \cdot \Delta t$, para $n=1, 2, \dots, N$ y Δt es la longitud constante del intervalo de muestreo. Como se puede apreciar, la primera diferencia notable entre el periodograma clásico (ec. 7) y el de Lomb-Scargle (ec. 1) es la forma en cómo se definen. Las otras diferencias se presentan a continuación.

1. **Frecuencias ortogonales.** Tradicionalmente, el periodograma clásico para series equiespaciadas en el tiempo está definido [33, 38] para un conjunto discreto de frecuencias ortogonales f_r , para las cuales las estimaciones del periodograma son independientes entre sí; esto es:

$$f_r = \frac{r}{T_{Tot}} \quad (8)$$

para $r = 0, 1, \dots, \lfloor N/2 \rfloor$, donde $\lfloor N/2 \rfloor$ es el número máximo de frecuencias y está definido por la función máximo entero, y donde $T_{Tot} = t_N - t_1$, i. e., la duración de la serie temporal y para el caso equiespaciado $T_{Tot} = (N - 1) \cdot \Delta t$. Las frecuencias f_r , al ser utilizadas en la relación (7), aseguran que el poder espectral correspondiente a diferentes frecuencias sea independiente para cualquier par de frecuencias del conjunto de frecuencias ortogonales, y que sea retenida la máxima cantidad de información posible de los datos originales [45]. Sin embargo, para el caso no equiespaciado en el tiempo, no es posible encontrar un conjunto de frecuencias ortogonales para las cuales el poder espectral sea independiente [38, 45].

2. **La frecuencia (razón) de muestreo y la frecuencia de Nyquist.** Para el caso equiespaciado en el tiempo, la frecuencia de muestreo f_s puede definirse como [32]

$$f_s = \frac{1}{\Delta t}, \quad (9)$$

por lo cual la frecuencia de Nyquist f_{Nyq} viene dada por

$$f_{Nyq} = \frac{f_s}{2} = \frac{1}{2\Delta t}. \quad (10)$$

Para el caso no equiespaciado en el tiempo, no hay definida una frecuencia de muestreo, y por tanto, una frecuencia de Nyquist [38, 45]. No obstante, hay varias propuestas para manejar este problema. Por ejemplo, Van-Dongen *et al.* [45] proponen utilizar la frecuencia real de Nyquist calculada por medio de la ventana espectral (*periodogram window*) (véase la sección 2.3 para más detalles). Bretthorst [4], en cambio, propone la frecuencia crítica de Nyquist, definida como

$$f_{cNyq} = \frac{1}{2\Delta t'}, \quad (11)$$

donde $\Delta t'$ es siempre menor o igual que el más pequeño intervalo de tiempo de muestreo. Por último, Schulz y Statteger [43] proponen una frecuencia de Nyquist promedio, esto es

$$\langle f_{Nyq} \rangle = \frac{1}{2 \langle \Delta t \rangle}, \quad (12)$$

donde $\langle \Delta t \rangle$ es el valor medio de los intervalos temporales de muestreo.

3. **El teorema de Parseval.** Una propiedad importante del periodograma clásico, para series temporales equiespaciadas en el tiempo, es el teorema de Parseval, i.e, la energía total de una señal real es la misma ya sea si es calculada en el dominio del tiempo o en el de la frecuencia [32]. Sin embargo, cuando las series no están equiespaciadas en el tiempo, este teorema no se cumple [45].

2.3 Estimación de las frecuencias naturales

Como se mencionó anteriormente, las frecuencias definidas por la relación (8), para el caso no equiespaciado, no son ortogonales. Sin embargo, de acuerdo con Fullerton [6], estas frecuencias pueden considerarse aproximadamente ortogonales para el caso no equiespaciado [6, 45]. A este conjunto de frecuencias utilizadas para calcular el periodograma se conocen como *frecuencias naturales*; están definidas [38, 45] por:

$$f'_r = \frac{r'}{T_{Tot}}, \quad (13)$$

para $r' = 0, 1, 2, \dots, M_{max}$, donde M_{max} es el número máximo de frecuencias naturales.

Para el caso en que los muestreos no están equiespaciados en el tiempo, M_{max} depende de los tiempos de muestreo, y la frecuencia $f_{M_{max}}$ no necesariamente corresponde con la frecuencia de Nyquist –que aparecería si las muestras estuvieran equiespaciadas temporalmente–. Por lo tanto, es necesario utilizar una alternativa válida a la frecuencia de Nyquist [38, 45]. Por ejemplo, si se opta por emplear

la frecuencia real de Nyquist sugerida por Van-Dongen *et al.* [45], el procedimiento es el siguiente.

Para poder determinar la frecuencia real de Nyquist f_{aNyq} , se hace uso de una función matemática conocida como *ventana espectral* o *función de respuesta espectral* [45]. Esta función depende sólo de los tiempos de muestreo, $\{t_n\}_{n=1}^N$ y de la diferencia de las frecuencias $\Delta f = f_a - f_b$, donde f_a y f_b son dos frecuencias consecutivas cualesquiera. Para el caso en que el muestreo está equiespaciado en el tiempo, existe una fórmula exacta para representar la ventana espectral; pero para el caso correspondiente a un muestreo no equiespaciado en el tiempo, no existe una fórmula exacta para su representación [38, 45]. Sin embargo, es posible utilizar la siguiente aproximación [17, 38, 45].

$$W(\Delta f) = \frac{1}{N^2} \left[\left(\sum_{n=1}^N \cos(2\pi\Delta f(t_n - \theta)) \right)^2 + \left(\sum_{n=1}^N \sen(2\pi\Delta f(t_n - \theta)) \right)^2 \right], \quad (14)$$

donde $\theta = (t_1 + t_N)/2$. Los valores de $W(\Delta f)$ están definidos entre 0 y 1 y es una función simétrica con respecto a Δf , *i. e.*, $W(-\Delta f) = W(\Delta f)$ [45].

Para determinar la f_{aNyq} se hace uso del fenómeno conocido como *aliasing*. Esto es, para dos frecuencias cuya suma o diferencia es un múltiplo entero de la frecuencia de muestreo, es posible construir una pareja de sinusoides que, después del muestreo, no son distinguibles una de la otra debido a que contienen las mismas muestras (Figura 1).

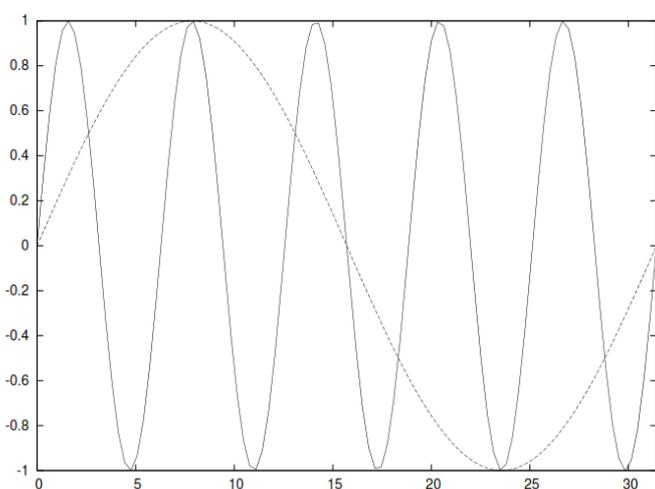


Figura 1. Dos diferentes sinusoides que se ajustan a un mismo conjunto de muestras y que ejemplifican el fenómeno del *aliasing*. Los tiempos (eje x) están en unidades arbitrarias. Elaboración propia.

Las copias debido al fenómeno del *aliasing* aparecerán cada $2mf_{aNyq}$, donde $m = \dots, -2, -1, 1, 2, \dots$. Por ejemplo, para el caso en que la serie temporal es cercanamente equiespaciada en el tiempo (las muestras están uniformemente distribuidas), se puede utilizar una f_{aNyq} , que equivale al caso en que la serie está equiespaciada. A continuación se realiza una inspección visual, para ello se dibuja la ventana espectral relación (14) utilizando como límite de las frecuencias a la f_{aNyq} y un valor de m mayor o igual a 2 ‘réplicas’. Cuando la ventana espectral se repite en un determinado valor de m , entonces el valor de f_{aNyq} es el límite para las frecuencias en el que se puede calcular el periodograma de Lomb-Scargle [15, 45].

En el caso de que las muestras no equiespaciadas temporalmente formen grupos que estén espaciados más o menos regularmente, entonces la máxima frecuencia natural estará determinada por el número de grupos n' [15, 45]; esto es:

$$f_{aNyq} = \frac{(n' - 1)/2}{T_{Tot}}. \quad (15)$$

Nótese que M_{max} no está determinado por el número de muestras. Las muestras individuales dentro de un grupo pueden ser consideradas como medidas redundantes obtenidas aproximadamente al mismo tiempo; por ello, más muestras dentro de cada grupo no incrementan la f_{aNyq} . Sin embargo, incrementan la significación estadística del periodograma [45].

3 El periodograma crudo vs. suavizado

El periodograma crudo, *i. e.*, el que se obtiene utilizando la relación (1), es un estimador inconsistente del espectro [3]. Esto es, el radio de la desviación estándar o coeficiente de variación del periodograma no tiende a cero a medida que el número de elementos de la serie aumenta [3, 23]. Un método para mejorar la estimación del espectro en términos de tendencia, varianza y supresión de falsos picos espectrales –de modo que se obtenga un estimador consistente– fue propuesto por Welch [49]. Esta técnica es conocida como *Welch-Overlapped-Segment-Averaging* (WOSA), y fue originalmente diseñada para el caso correspondiente a series temporales equiespaciadas en el tiempo [49]. Para el caso de series no equiespaciadas temporalmente, Schulz y Statteger [43] hicieron una adaptación de la técnica de Welch. El método propuesto por estos autores [43], consiste en partir la serie temporal que se va a estudiar (de longitud N) en varios segmentos (n_{50}) de longitud $N_{Seg} = 2N/(n_{50} + 1)$, de tal modo que cada uno de estos segmentos se solape durante la mitad de su longitud (véase Figuras 2 y 3).

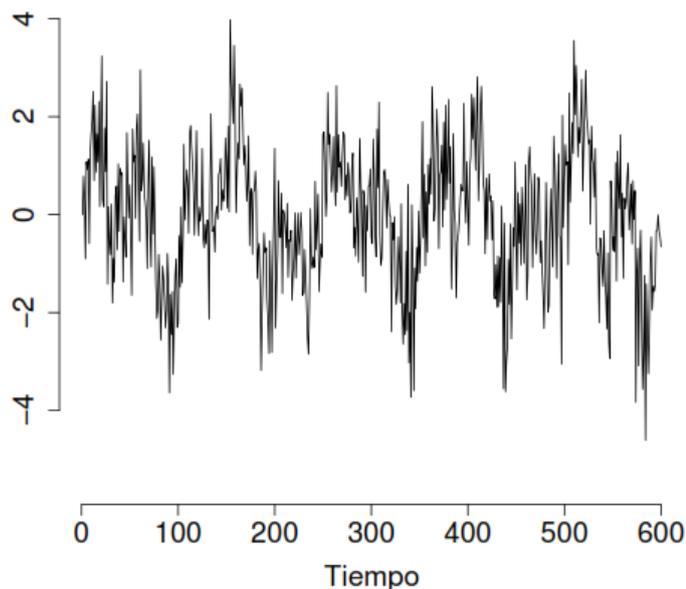


Figura 2. Serie temporal ejemplo (suma de dos funciones armónicas con periodos $2\pi/50$ y $2\pi/120$ más ruido Gaussiano $\mathcal{N}(0, 0.95)$) con N elementos a la cual se le aplica la técnica WOSA. Elaboración propia.

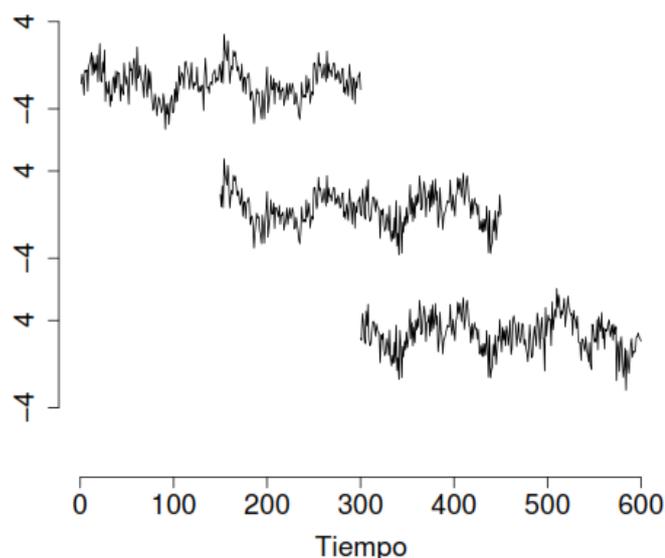


Figura 3. Diagrama donde se muestra el procedimiento de la técnica WOSA, utilizando la serie de la Figura 2, con $n_{50} = 3$ y $N_{seg} = N/2$. Elaboración propia.

En el espectro, para evitar las irregularidades de los extremos de los N_{seg} segmentos, al usar WOSA, se ha de aplicar (multiplicar) algún tipo de función-ventana (conocida como *taper*) a cada segmento n_{50} en el dominio del tiempo [43, 49]. Existen diferentes tipos de ventanas (una amplia revisión puede encontrarse en [12, 26]), pero una de las más utilizadas es la ventana de Hann o de Hanning, la cual está definida [12] como:

$$w(n) = 0.5 \left[1 - \cos \left(\frac{2\pi n}{N_{seg} - 1} \right) \right], \quad (16)$$

para $n=0, 1, \dots, N_{seg} - 1$.

La función-ventana es escalada de esta forma $\sum w_n^2 = N_{seg}$. A continuación, para los n_{50} segmentos de la serie temporal a los cuales se les ha aplicado la función ventana escalada, se calcula el periodograma de Lomb-Scargle mediante (1). Por último, se calcula el promedio de los n_{50} espectros crudos. El espectro promedio resultante es un estimador consistente del (auto) espectro [3, 43]; se le conoce como espectro suavizado y se define de este modo, de acuerdo con Schulz y Stattegger [43]

$$\hat{G}_{xx}(f_i) = \frac{2}{n_{50}\Delta f N_{seg}} \sum_{n=1}^{n_{50}} |X_n(f_i)|^2, \quad i = 1, 2, \dots, k, \quad (17)$$

donde $\Delta f = 1/(N_{seg} \langle \Delta t \rangle)$ y $|X_n(f_i)|^2$ es el periodograma de Lomb-Scargle definido en la relación (1).

Dado el escalado del espectro suavizado relación (17), se cumple la siguiente relación con la varianza de la serie temporal:

$$\Delta f \sum \hat{G}_{xx} = \sigma_x^2, \quad (18)$$

donde Δf es la frecuencia fundamental [43].

Debido a que las componentes de los espectros crudos siguen una distribución del tipo χ^2 [30, 33] con 2 grados de libertad, (17) también sigue ese tipo de distribución [43]. Cada uno de los n_{50} espectros de la ecuación (17) incrementa los grados de libertad, por lo cual se reduce el error estándar de la estimación del espectro. Sin embargo, el solapamiento de los n_{50} segmentos introduce una correlación entre los segmentos; por ello, es necesario introducir el concepto de número efectivo de segmentos n_{eff} , donde el n_{eff} es mucho menor que el n_{50} , y la relación entre ellos de acuerdo con Schulz y Stattegger [43], viene dada por

$$n_{eff} = n_{50} \left(1 + 2c_{50}^2 - \frac{2c_{50}^2}{n_{50}} \right)^{-1}, \quad (19)$$

donde c_{50} es una constante que depende del tipo de la función ventana, por ejemplo, para la ventana de Hanning $c_{50} = 0.167$ [12]. Por tanto, los grados de libertad ν para (17) dependerán del n_{eff} relación (19), y por tanto, de la constante c_{50} ; esto es:

$$\nu = 2n_{eff} \quad (20)$$

La distribución del auto-espectro (17) viene dada por la relación [30, 33]

$$\frac{\hat{G}_{xx}(f_i)}{G_{xx}(f_i)} = \frac{\chi^2}{\nu}. \quad (21)$$

De aquí, se sigue que un $(1 - \alpha)$ intervalo de confianza para $G_{xx}(f_i)$, basado en la estimación de $\hat{G}_{xx}(f_i)$, puede ser calculado por [3]

$$\left[\frac{\nu \hat{G}_{xx}(f_i)}{\chi^2_{\nu, \alpha/2}} \leq G_{xx}(f_i) \leq \frac{\nu \hat{G}_{xx}(f_i)}{\chi^2_{\nu, 1-\alpha/2}} \right]. \quad (22)$$

Algunos autores [3, 30, 33] aplican una transformación logarítmica del tipo $10 \log_{10}$ a la relación (22), esto con el objeto de que los intervalos de confianza sean independientes de la frecuencia y al graficar el \hat{G}_{xx} se obtenga una mejor visualización.

La mínima resolución de las frecuencias en \hat{G}_{xx} , *i. e.*, el ancho de banda espectral, B_w , se calcula como:

$$B_w = \beta_w \Delta f, \quad (23)$$

donde β_w es el ancho de banda normalizado y depende del tipo de ventana (*taper*) utilizada; por ejemplo, para la ventana de Hanning tiene un valor de 2 [12, 26, 43].

4 Contraste de hipótesis del periodograma de Lomb-Scargle debido a ruido rojo

Es bien sabido que muchos fenómenos ambientales (climáticos, meteorológicos, etc.) o geofísicos presentan un espectro de fondo rojo, *i. e.*, la intensidad o amplitud espectral disminuye exponencialmente a medida que decrecen sus frecuencias [10, 13, 42]. [13] propuso que un proceso estocástico autoregresivo de primer orden (AR1) es un modelo adecuado para representar ese espectro de fondo rojo, y que puede ser utilizado como una hipótesis nula en un test de significación [23, 42]. Si $R(t_r)$ es un proceso AR1, donde los tiempos de muestreo t_r ($r = 1, 2, \dots, N$) no están equiespaciados, la distribución del periodograma de Lomb-Scargle para este proceso no puede ser calculada analíticamente [23, 24, 42]. Por tanto, es necesario utilizar métodos de simulación de Monte-Carlo para inferir las propiedades estadísticas de la distribución del periodograma [24]. El sesgo en el espectro es debido a que las componentes individuales del periodograma de Lomb-Scargle no son independientes unas de las otras [17, 38].

Schulz y Mudelsee [42] demostraron, por medio de simulaciones de Monte Carlo, que el periodograma de Lomb-Scargle de un proceso AR1 no equiespaciado temporalmente sobrestima las amplitudes espectrales localizadas en las altas frecuencias (*i. e.*, las frecuencias al final del espectro). Ellos también desarrollaron un algoritmo computacional, conocido como REDFIT, para corregir ese sesgo. De hecho, hasta donde se sabe, es el único algoritmo y programa computacional existente que tiene en cuenta el ruido rojo de fondo en la estimación del espectro suavizado de series temporales ambientales no equiespaciadas, vía el periodograma de Lomb-Scargle.

A continuación, se presentan los principales conceptos estadísticos para luego proporcionar la descripción de REDFIT, siguiendo principalmente a Mudelsee [23] y Schulz y Mudelsee [42]. Más detalles pueden consultarse en [22, 23, 24, 42].

Un proceso discreto autoregresivo de orden 1 (AR1) y no equiespaciado temporalmente está definido [24, 36, 42] por la siguiente relación:

$$\begin{aligned} R(t_1) &= \mathcal{E}_{N(0,1)}, & \text{para } r=1 \\ R(t_r) &= \rho_r R(t_{r-1}) + \mathcal{E}_{N(0,\sigma_{\mathcal{E}})}(t_r), & \text{para } r=2, 3, \dots, N \\ \rho_r &= \exp(-(t_r - t_{r-1})/\tau_m), \end{aligned} \quad (24)$$

donde τ_m es la escala temporal correspondiente al proceso AR1 (una medida de su memoria o persistencia) y \mathcal{E} es un ruido Gaussiano de media cero y varianza

$$\sigma_{\mathcal{E}}^2 = 1 - \exp(-2(t_r - t_{r-1})/\tau_m). \quad (25)$$

El valor de $\sigma_{\mathcal{E}}^2$ asegura que el proceso AR1 sea estacionario y tenga varianza unitaria [42].

El espectro teórico² $G_{rr}(f_i)$ de un proceso AR1 (ecuación 24) está definido [30, 42] por

$$G_{rr}(f_i) = \frac{1 - \langle \rho \rangle^2}{G_0 [1 - 2 \langle \rho \rangle \cos(\pi f_i / \langle f_{Nyq} \rangle) + \langle \rho \rangle^2]}, \quad (26)$$

para las frecuencias $f_i, j = 1, 2, \dots, K$, desde la frecuencia fundamental Δf (véase la relación (17)) hasta la frecuencia promedio de Nyquist [42]. Allí, G_0 es la amplitud espectral promedio, $\langle \rho \rangle$ es el coeficiente de autocorrelación promedio; *i. e.*, $\langle \rho \rangle = \exp(-\langle \Delta t \rangle / \tau_m)$, donde $\langle \Delta t \rangle$ es la media de los intervalos temporales del proceso AR1 [42] y el valor de τ_m se estima mediante el algoritmo y programa computacional de nombre TAUEST, propuesto por Mudelsee [22].

Con el periodograma de Lomb-Scargle, la desviación sistemática entre un espectro teórico de ruido rojo (relación 26) y otro espectro estimado dependerá de los tiempos de muestreo [17, 38]. Pero la carencia de una solución analítica para la desviación (sesgo) impide aplicar una corrección directa al periodograma de Lomb-Scargle [42]. Para superar este obstáculo, se utilizan métodos de Monte-Carlo en combinación con técnicas de *bootstrap*. Basada en los tiempos de muestreo, se genera un *ensemble* de N series temporales sintéticas AR1 (Ecuación (24)). La desviación del espectro promedio del *ensemble* con respecto al espectro teórico (Ecuación (26)) es utilizada para hacer la corrección [24, 42].

A continuación, se describe el algoritmo REDFIT tal como es presentado por Schulz y Mudelsee [42].

²Para el caso equiespaciado, el término coseno del espectro teórico de un AR1 está definido [16] por $\cos(2\pi f_i \Delta t)$, para $-\frac{1}{2\Delta t} \leq f_i \leq \frac{1}{2\Delta t}$. Por esta razón, el término coseno en la relación 26 al sustituir $2 \langle \Delta t \rangle$ por $\frac{1}{\langle f_{Nyq} \rangle}$ se expresa como $\cos(\pi f_i / \langle f_{Nyq} \rangle)$.

1. Se estima τ_m para la serie temporal a estudio $X(t_r)$ ($r = 1, 2, \dots, N$) utilizando el algoritmo TAUEST de Mudelsee [22]. En caso de utilizar varios n_{50} segmentos WOSA, se calcula el promedio de los τ_m 's de todos los n_{50} segmentos.
2. Se estima, mediante el PLS, el espectro $\widehat{G}_{xx}(f_i)$ de $X(t_r)$ en el intervalo $[0, < f_{Nyq} >]$, y se determina el área bajo $\widehat{G}_{xx}(f_i)$ para estimar la varianza de $X(t_r)$.
3. Se realizan $p=1$ a N_{sims} simulaciones de Monte-Carlo.
 - Se crea la p -ésima serie temporal AR1 por medio de (24), usando los tiempos t_r 's de la serie $X(t_r)$, el τ_m estimado en el paso 1 y un conjunto independiente de $\mathcal{E}(t_r)$'s.
 - Se estima el espectro $\widehat{G}_{rr}(f_i)$ para la serie AR1.
 - Se escala $\widehat{G}_{rr}(f_i)$ de tal modo que su área bajo el espectro sea igual al área bajo el espectro de $\widehat{G}_{xx}(f_i)$.

Se determina la media aritmética de los N_{sims} espectros correspondientes a las series AR1, $< \widehat{G}_{rr}(f_i) >$

4. Se calcula el espectro teórico $G_{rr}(f_i)$ de un proceso AR1, utilizando la relación (26) y el valor estimado de τ_m .
5. Se calcula G_o (relación 26) de tal modo que el área bajo $G_{rr}(f_i)$ sea igual al área bajo $\widehat{G}_{xx}(f_i)$.
6. Se calcula un factor de corrección $c(f_i)$ para ajustar la desviación del espectro Lomb-Scargle, *i. e.*

$$c(f_i) = \frac{\langle \widehat{G}_{rr}(f_i) \rangle}{G_{rr}(f_i)}. \quad (27)$$

7. Se calcula el espectro de Lomb-Scargle corregido:

$$\widehat{G}'_{xx}(f_i) = \frac{\widehat{G}_{xx}(f_i)}{c(f_i)}. \quad (28)$$

8. Para calcular la significación estadística de un pico espectral, el intervalo superior de confianza del ruido AR1 es calculado para varios niveles de significación (*v. gr.*, 0.95 o 0.99), siguiendo una distribución del tipo χ^2 con ν grados de libertad (calculados, con (20)). También es posible calcular niveles de significación a partir de los percentiles del *ensemble* de Monte-Carlo.
9. Se verifica si el modelo AR1 es adecuado para describir $X(t_r)$ probando la igualdad de $G_{rr}(f_i)$ y $\widehat{G}'_{xx}(f_i)$ usando el test no paramétrico de Wald-Wolfowitz más comúnmente conocido como *runs test* [3].

³<ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/greenland/summit/gisp2/isotopes/gispd18o.txt>

⁴La serie alcanza los 110.000 años antes del presente, pero es muy factible que los errores de datación para estas edades pudieran afectar la estimación del espectro.

5 Algunas consideraciones prácticas en la estimación del espectro

En esta sección, se presentan las principales consideraciones prácticas que se deben tener en cuenta para la estimación del espectro suavizado vía el periodograma de Lomb-Scargle; ello se realiza mediante dos ejemplos con series temporales no equiespaciadas temporalmente. Se hace énfasis en cómo utilizar esta metodología y en la interpretación de los resultados. El primer ejemplo corresponde a una serie temporal paleoclimática utilizada para estimar temperaturas del aire en la superficie de la Tierra; proviene de testigos de hielo de Groenlandia y cubre los últimos 50.000 años [11]. El segundo ejemplo corresponde a una serie temporal de capturas de atún rojo obtenidas en las almadrabas del duque de Medina Sidonia, las cuales estuvieron operativas durante el periodo 1525 a 1816 [18]. Para ambos ejemplos se utilizó el paquete computacional de uso gratuito REDFIT (ver. 3.5) creado por Schulz y Mudelsee [42], y que se encuentra disponible en <http://www.ncdc.noaa.gov/paleo/softlib/redfit/redfit.html>.

5.1 Caso práctico 1: valores medios de $\delta^{18}\text{O}$ de GISP2

La serie temporal no equiespaciada utilizada en este ejemplo corresponde a valores medios de $\delta^{18}\text{O}$ (per mil); proviene del proyecto de testigos de hielo GISP2 (*the Greenland Ice Sheet Project Two*) y fue obtenida del *National Climatic Data Center*, NOAA³. El principal interés científico de ésta serie se debe al hecho de que ha sido utilizada como un buen paleo-indicador de temperaturas superficiales del aire para, al menos, el área geográfica a la que pertenece [11]. En este trabajo, se hace énfasis en el intervalo 15 a 50.000 años⁴ (Figura 4), por dos razones; primero, debido a que *a priori* se conoce que presenta un ciclo de *ca.* 1470 años [11, 40]; segundo, porque el intervalo 15-50.000 años presenta una gran irregularidad en el muestreo (con un valor medio de 110 años) (Figura 5), en comparación con los últimos 10.000 años (con un valor medio de 13 años); finalmente, debido a que las técnicas de análisis espectral requieren que la series a estudio sean estacionarias y, como se puede observar (Figura 4), la serie de $\delta^{18}\text{O}$ que cubre los 50.000 años no parece satisfacer este requerimiento. Una posible solución para sortear este inconveniente es dividir la serie en dos partes con características estadísticas similares. Esta división, en este caso de estudio, también está sustentada desde el punto de vista de la física del clima, debido a que el segmento temporal del Holoceno (últimos 11.000 años) presenta características climáticas diferentes respecto del periodo inter-glaciar (entre 12 y 50.000 años).

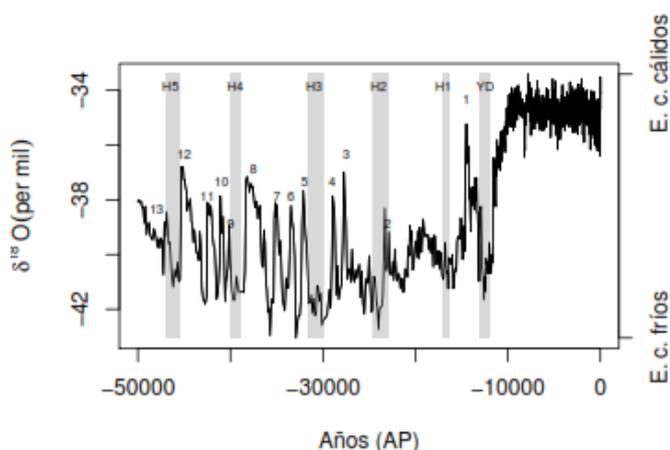


Figura 4. Serie temporal de valores medios del isótopo estable $\delta^{18}\text{O}$ (per mil) de GISP2 [11] para los últimos 50.000 años (donde AP indica años antes del presente). Las áreas rectangulares en gris representan los eventos Heinrich (H1-H5) y el Younger-Dryas (YD). Los números del 1-13 son los ciclos Dansgaard-Oeschger (DO). El lado derecho de las ordenadas representa los eventos climáticos cálidos y fríos. Elaboración propia.

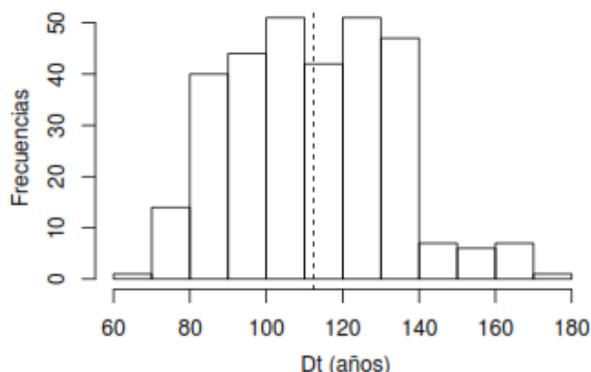


Figura 5. Histograma de las diferencias (distancias) entre los tiempos de muestreo (Dt) de la serie de $\delta^{18}\text{O}$ de GISP2 para el intervalo 14 a 50.000 años. La línea vertical discontinua es el valor medio de las diferencias entre los tiempos de muestreo. Elaboración propia.

Antes de llevar a cabo la estimación espectral para la serie temporal de los valores de $\delta^{18}\text{O}$, se verifica si el espectro de la serie a estudio es consistente con el modelo de ruido rojo utilizado [42]. Para ese fin, se utiliza el test Wald-Wolfowitz (conocido también como *runs test*) [3], el cual está incorporado en el paquete REDFIT [42]. El valor del estadístico para este caso es de 79, y está dentro del intervalo de confianza al 95 %, por lo cual se puede aceptar que el modelo AR1 utilizado es adecuado para caracterizar el espectro de la serie de $\delta^{18}\text{O}$. La estimación del espectro suavizado de la serie temporal después de remover la tendencia lineal de $\delta^{18}\text{O}$ se muestra en la Figura 6, mientras que los parámetros de entrada para generar el espectro vía REDFIT se presenta en la Tabla 1.

Entre estos parámetros de entrada destaca el número de segmentos WOSA (N50) debido a que está relacionado di-

rectamente en el suavizado del periodograma. Un valor muy grande del número de segmentos WOSA determina un periodograma demasiado suavizado, por lo cual se podría difuminar algún tipo de señal contenida en la serie temporal a estudio. Por el contrario, si se utiliza un solo segmento, se obtendrá un periodograma "crudo" (con dos grados de libertad); y es bien sabido que este periodograma no es un estimador consistente del espectro, y que puede contener picos espectrales que podrían no representar eventos periódicos reales [3, 23]. Debido a que no existe una regla precisa para determinar el número de segmentos WOSA, es recomendable que el espectro suavizado tenga al menos 4 ó 5 grados de libertad [48]. Una estrategia muy utilizada para la determinación de N50 consiste en calcular el espectro de la serie bajo análisis con diferentes números de segmentos WOSA, y así decidir cuál es el N50 más apropiado.

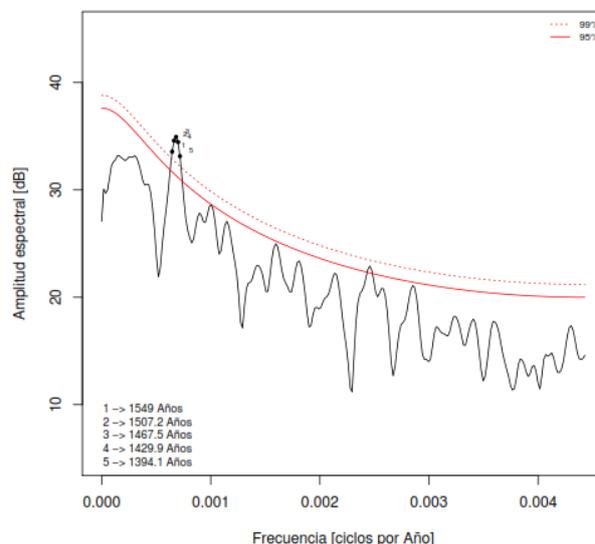


Figura 6. Espectro corregido y suavizado (con 7 grados de libertad) (curva continua en negro) de la serie temporal de $\delta^{18}\text{O}$ de GISP2, para el intervalo 14 a 50.000 años. Las curvas discontinuas en rojo representan el nivel del ruido rojo con un 95 % y 99 % de confianza de una distribución χ -cuadrada para un proceso AR1. Elaboración propia.

Como se puede apreciar (Figura 6), el pico espectral más destacado está localizado *ca* 1470 años, el cual supera el espectro asociado a un ruido de tipo AR1 con un nivel de confianza superior del 95 %. Este pico corresponde a la conocida característica espectral de $\delta^{18}\text{O}$ de GISP2, y se sabe que se presenta con más intensidad en el intervalo de 11-74.000 años (AP) [40]. Esta periodicidad *ca* de 1470 años está relacionada con los eventos Dansgaard-Oeschger (DO), *i.e.*, fluctuaciones rápidas del clima entre eventos climáticos fríos y cálidos (Figura 4), con una ocurrencia media de 1470 años y que tuvieron lugar durante el último periodo glacial (entre unos 11 a 100.000 años AP); no obstante, hoy por hoy no existe una hipótesis única para explicar su origen -más allá de que está fuera de los objetivos de este trabajo el profundizar en dichas razones-.

| Parámetro | Valor |
|------------------------------------|------------------|
| No. de simulaciones de Monte-Carlo | Nsim = 5000 |
| Test de Monte-Carlo | mctest = T |
| Oversampling factor | ofac = 4.0 |
| No. de segmentos WOSA | N50 = 4 |
| Tipo de ventana | iwin = 1 (Welch) |

Tabla 1. Parámetros de entrada para REDFIT para la estimación del espectro de $\delta^{18}\text{O}$ de GISP2. Elaboración propia.

5.2 Caso práctico 2: capturas artesanales de atún rojo

El segundo ejemplo, en este trabajo, corresponde a una serie temporal de capturas anuales de atún rojo realizadas durante el periodo 1525-1816 (Figura 7). Estos datos provienen de López-Capont [18] aunque, para este trabajo, fueron obtenidos de Ganzedo *et al* [7] y, principalmente de López-González y Ruiz (Tabla 1) [19]. Son capturas anuales en número de individuos, capturas provenientes de cada una de las almadrabas del duque de Medina Sidonia, principalmente de Conil y Zahara; ellas estuvieron operativas durante todo el periodo de estudio (1525-1816). La serie cubre 292 años, con 21 datos perdidos [7, 19, 31], estas pérdidas se deben a que en ciertos años no se hicieron los registros de las capturas o no se armaron las almadrabas [18, 19].

A primer golpe de vista, la serie de las capturas (Figura 7) no parece ser estacionaria, por la tendencia decreciente del número de capturas en el transcurso del tiempo.

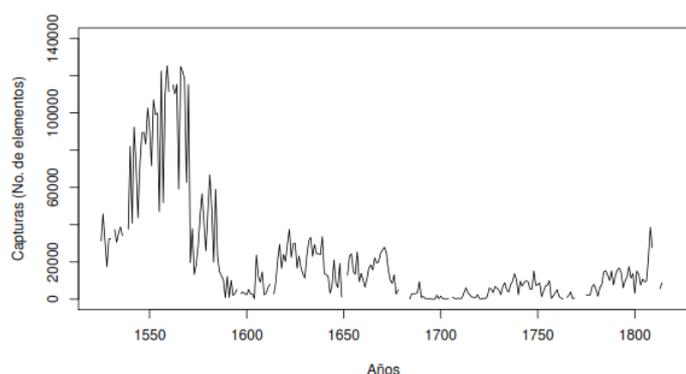


Figura 7. Serie temporal de capturas artesanales de atún rojo de las almadrabas del duque de Medina-Sidonia para el periodo 1525-1816 [7, 18, 19]. Elaboración propia.

La opción de dividir la serie en dos trozos con características estadísticas parecidas está limitada en cierta medida debido al número de elementos de cada trozo, de tal manera que la primera parte de la serie abarcaría unos 70 (elementos) años, mientras que la segunda parte, unos 220 años. Sin embargo, teniendo en cuenta estas limitaciones, se estiman sus periodogramas suavizados de los dos trozos, así como de toda la serie; esto, con el objeto de utilizar

toda la información disponible para averiguar potenciales eventos periódicos registrados en las series de las capturas. Nótese que, a diferencia de la serie temporal de valores de $\delta^{18}\text{O}$ de GISP2, que se conocía a priori sus características espectrales, en la serie de las capturas no se cuenta con esa información.

Como se observa en los periodogramas suavizados (Figura 8), se encontraron varios picos espectrales que superan el umbral de ruido con un nivel de confianza del 95%. El primero (Figura 8, superior derecha), de unos 78 años, aparece en el periodograma suavizado de toda la serie temporal de las capturas; en contraste, no parece evidente la presencia de este pico espectral en el periodograma del intervalo 1600-1816, según se observa. Una de las cuestiones más importantes que se debe tener en cuenta al aplicar cualquier técnica de análisis espectral, es que los picos en las primeras frecuencias (de izquierda a derecha) pueden contener parte de la tendencia de las series bajo análisis [48]. Es cierto que la tendencia lineal ha sido removida antes de calcular el espectro, pero también hay que tener en cuenta que el cálculo de la tendencia en una serie temporal es algo complicado, y es muy difícil separar tendencia, periodicidades de muy baja frecuencia y ruido [45, 50] de forma suficientemente precisa. Por ello, es difícil sustentar que el pico espectral de *ca.* 80 años pueda estar relacionado con algún evento periódico; para ello se necesitaría que la serie cubriera un intervalo temporal mucho más amplio.

El segundo pico espectral (Figura 8, superior izquierda) estadísticamente significativo está localizado alrededor de los 20 años, aunque este pico sólo está presente en el espectro del intervalo 1525-1600. Sin embargo, téngase en cuenta que un ciclo con un período de 20 años solo cabe poco más de tres veces en el intervalo 1525-1600. Por otro lado, este pico espectral de *ca.* 20 años no está presente en el espectro del intervalo 1600-1816, ni en toda la serie (intervalo 1525-1816). Por estas razones, es muy probable que no represente una señal periódica real.

Sin lugar a dudas, el resultado más interesante en este caso de estudio, son los picos espectrales estadísticamente significativos en las altas frecuencias, que podrían representar eventos periódicos reales entre 2.3 años y 3.5 años. De entre estos picos, el que está localizado alrededor de 3.5 años, se halla presente en los tres periodogramas (Figura 8), lo cual le añade mayor evidencia estadística. Antes de presentar una discusión sobre las posibles causas de los ciclos de 2.3 y 3.5 años, es importante tener en cuenta que las frecuencias de estos picos espectrales están relativamente lejanas entre sí (superan varias veces el ancho de banda en cada uno de los espectros), lo cual podría representar que existieron eventos periódicos independientes. Los ciclos de 2.3 y 3.5 años podrían estar relacionados con forzamientos ambientales que presentan una fuerte variabilidad de tipo periódico o cuasi-periódico cercana a esas frecuencias. Por ejemplo, es bien sabido que la Oscilación del Atlántico Norte (OAN) presenta en su espectro picos de corto periodo, de entre 2.2 y 2.4 años. Por otro lado, el fenómeno

del Niño-Oscilación del Sur (ENOS) presenta un gran poder espectral de corto periodo, de alrededor de 3.5 años [8]. Sin embargo, tampoco se descarta que estos eventos periódicos están relacionados con factores intrínsecos relacionados con la dinámica poblacional del atún rojo [7, 31]. Una discusión a detalle queda fuera del objetivo de este trabajo.

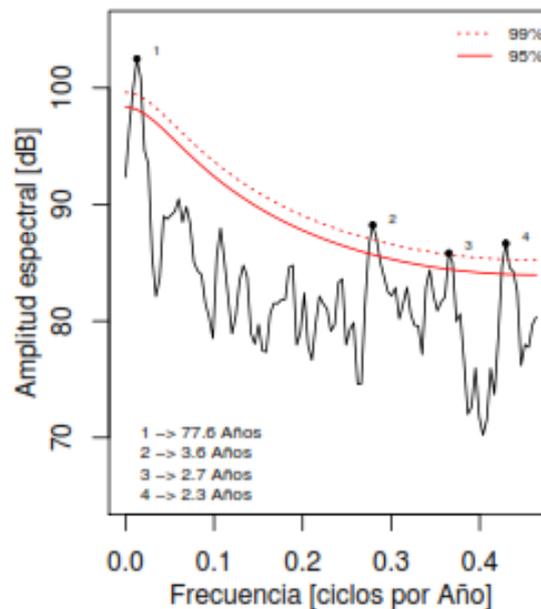
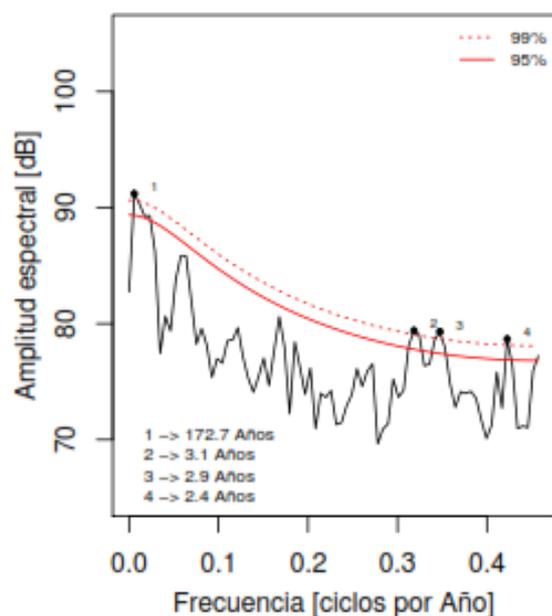
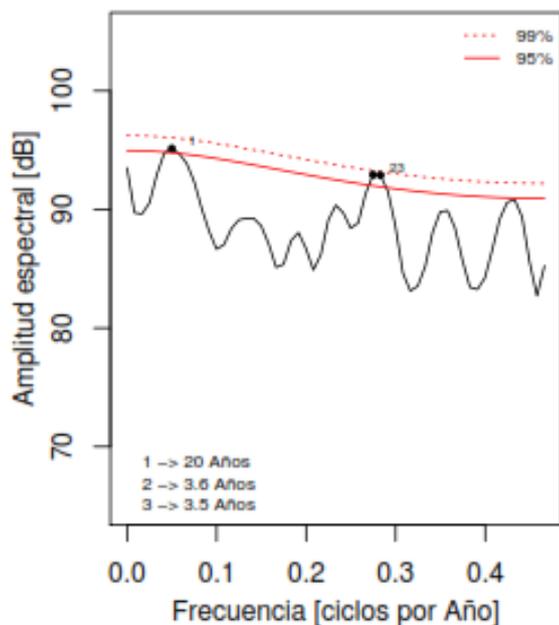


Figura 8. Espectros corregidos y suavizados (con 5, 6 y 7 grados de libertad, respectivamente) (curvas continuas en negro) de los segmentos que cubren los periodos 1525-1600 (superior izquierda), 1600-1816 (inferior izquierda) y todo el periodo 1525-1816 (superior derecha) de las capturas de atún rojo de las almadrabas de Medina-Sidonia [18]. Las curvas en rojo representan el nivel del ruido rojo con un 95% y 99% de confianza de una distribución χ -cuadrado para un proceso AR1. Elaboración propia.

6 Conclusiones

El análisis espectral es una de las técnicas estadísticas más utilizadas para el análisis de series temporales ambientales, ya sea para la búsqueda e identificación de eventos periódicos o para estudiar el comportamiento de una serie temporal en el dominio de la frecuencia. Cuando las series temporales ambientales están equiespaciadas, existen varias técnicas de análisis espectral para estimar el espectro, así como suficiente documentación sobre ello. Sin embargo, cuando las series no están equiespaciadas, estimar el espectro no suele ser una tarea fácil. Esto es debido a que no existe una gran diversidad de técnicas de análisis espectral que puedan ser utilizadas de modo directo con series temporales ambientales no equiespaciadas y, como consecuencia, hay poca documentación al respecto. Una de estas técnicas de análisis espectral es el periodograma de Lomb-Scargle [17, 38, 39], originalmente diseñado para analizar series temporales astronómicas no equiespaciadas temporalmente; dicha técnica fue adaptada por Schulz y Stattegger [43] para trabajar con datos ambientales y mejorada por Schulz y Mudelsee [42] para tener en cuenta el tipo de ruido de fondo de las series temporales ambientales.

En lo que respecta a la disponibilidad de información de la adaptación del periodograma de Lomb-Scargle para analizar series temporales ambientales teniendo en cuenta ruido de fondo rojo, en lengua inglesa están los trabajos pioneros

de Schulz y Stattegger [43], Mudelsee [22], Schulz y Mudelsee [42], Mudelsee *et. al.*, [24], y muy bien resumidos en Mudelsee [23]. Sin embargo, en idioma español existe muy poca información al respecto (algunas excepciones son Polanco-Martínez [31] y Pardo-Igúzquiza y Rodríguez-Tovar [27], aunque ésta última no utiliza un ruido de fondo rojo); hasta donde se sabe, no se ha publicado trabajo alguno de revisión que englobe todo lo que se sabe de esta técnica. Por esta razón, en este artículo, se presentó el estado del conocimiento de la estimación espectral de series temporales ambientales no equiespaciadas temporalmente mediante el periodograma suavizado de Lomb-Scargle y teniendo en cuenta un tipo de ruido de fondo rojo.

En este trabajo, también se presentaron dos casos de estudio para ejemplificar de forma práctica la estimación del periodograma suavizado de Lomb-Scargle teniendo en cuenta el ruido rojo de fondo de las series temporales ambientales. Para tal fin, se usó el paquete computacional REDFIT [42]. En el primer caso de estudio, se utilizó una serie temporal no equiespaciada (con mucha irregularidad) de valores de $\delta^{18}\text{O}$, provenientes del mundialmente reconocido proyecto GISP2. Como se pudo apreciar, se encontró un pico espectral estadísticamente significativo al 95% *ca.* 1470 años, el cual es la conocida marca espectral de esta serie temporal. En el otro ejemplo, en los espectros de las capturas de atún rojo, se obtuvo unos picos significativos al 95% entre 2.4 y 3.5 años, los cuales podrían estar relacionados con los eventos de la OAN y el ENOS. Finalmente (anexo A), se proporcionó una breve descripción del paquete computacional REDFIT y un sencillo programa computacional, de fácil y libre modificación en lenguaje R (Anexo B) para graficar las salidas de REDFIT.

AGRADECIMIENTOS

Se agradece al Prof. J. Sáenz quién me inició en el mundo del análisis espectral de datos climáticos no equiespaciados temporalmente, así como, por la discusión de una parte del contenido de este artículo. Del mismo modo, se agradece al Dr. U. Gancedo por la discusión en la interpretación de varios aspectos relacionados con las series de capturas de atún rojo aquí presentadas. Asimismo, al Dr. M. Mudelsee por proporcionarme el código fuente de TAUEST, y en conjunto con el Prof. Dr. M. Schulz por la ayuda prestada para entender ciertas cuestiones técnicas del código fuente de REDFIT; de modo similar, por ofrecer de forma gratuita este paquete en internet. Un agradecimiento especial a los revisores (Analítika) de este artículo, así como también a Martha Rodríguez por sus atinados comentarios para la mejora de este trabajo. Un especial agradecimiento a la Lic. M. Arrugaeta por pulir el inglés del *abstract*. Este trabajo ha sido posible gracias al apoyo del proyecto FP7 BASE (*Button up Adaptation Strategies for a Sustainable Europe*) No. de Ref. 308337.

Referencias

- [1] I. AROSTEGUI, I. BARRIO, M. LEGARRETA and A. URKAREGI, *Opinión-contribución (J. M. Polanco M.) sec. 3.7 "técnicas estadísticas no estándar para el análisis de series temporales ambientales"* in "statistics in research vs. research in statistics", Boletín de Estadística e Investigación Operativa, 30 (2014), pp. 233–235.
- [2] F. J. M. BARNING, *The numerical analysis of the light-curve of 12 Lacertae*, Bull. of the Astron. Inst. of the Neth., 17 (1963), pp. 22–28.
- [3] J. S. BENDAT and A. G. PIERSOL, *Random Data. Anal. and Meas. Proc.*, Wiley-Interscience, New York, (1986).
- [4] G. BRETTHORST, *Frequency Estimation and Generalized Lomb-Scargle Periodograms in Statistical Challenges in Astronomy*, Edited by E. D. Feigelson and G. Jogesh Babu, Springer, New York., (2003).
- [5] G. FOSTER, *The cleanest Fourier spectrum*, The Astronomical Journal, 109 (1995).
- [6] A. W. FULLERTON, *Searching for periodicity in astronomical data.*, in The Study of Variable Stars Using Small Telescopes, Cambridge University Press, Cambridge., (1986).
- [7] U. GANCEDO, E. ZORITA, A. P. SOLARI, G. CHUST, A. S. DEL PINO, **Polanco, J.** and J. J. CASTRO, *What drove tuna catches between 1525 and 1756 in southern Europe?.*, ICES J. of Mar. Sc., (2009).
- [8] N. GARCÍA, L. GIMENO, L. DE LA TORRE, R. NIETO, and J. AÑEL, *North atlantic oscillation (NAO) and precipitation in galicia (Spain)*, *Atmósfera*, 18 (2005), pp. 25–32.
- [9] M. GHIL, M. R. ALLEN, M. D. DETTINGER, K. IDE, D. KONDRASHOV, M. E. MANN, A. W. ROBERTSON, A. SAUNDERS, Y. TIAN, F. VARADI and P. YIOU, *Advanced spectral methods for climatic time series*, *Rev. Geophys.*, 40 (2002), pp. 1003, doi:10.1029/2000RG000092.
- [10] D. L. GILMAN, F. J. FUGLISTER and J. M. MITCHELL JR, *On the power spectrum of "red noise"*, *J. Atm. Sc.*, 20 (1963), pp. 182–184.
- [11] P. M. GROOTES AND M. STUIVER, *Oxygen 18/16 variability in greenland snow and ice with 10- 3-to 105-year time resolution*, *Journal of Geophysical Research: Oceans*, 102 (1997), pp. 26455–26470.
- [12] F. J. HARRIS, *On the use of windows for harmonic analysis with the discrete Fourier transform*, *Proc. of the IEEE*, 66 (1978), pp. 51–83.
- [13] K. HASSELMANN, *Stochastic climate models Part I. Theory*, *Tellus*, 28 (1976), pp. 473–485.

- [14] D. HESLOP AND M. J. DEKKERS, *Spectral analysis of unevenly spaced climatic time series using CLEAN: signal recovery and derivation of significance levels using a Monte Carlo simulation*, Phys. Earth Planet. Int., 130 (2002), pp. 103–116.
- [15] J. H. HORNE and S. L. BALIUNAS, *A prescription for period analysis of unevenly sampled time series*, Astrophys. J., 302 (1986), pp. 757–763.
- [16] G. M. JENKINS and D. G. WATTS, *Spectral analysis and its applications*, Holden-Day, Oakland, CA, (1969).
- [17] N. R. LOMB, *Least-squares frequency analysis of unequally spaced data*, Astrop. Spa. Sc., 39 (1976), pp. 447–462.
- [18] F. LÓPEZ-CAPONTE, *La etapa pesquera del padre Sarmiento y su época: de los atunes, y sus transmigraciones y conjeturas, sobre la decadencia de las almadrabas, y sobre los medios para restituirlas.*, Caixa de Pontevedra, Pontevedra, (1997).
- [19] J. LÓPEZ GONZÁLEZ and J. RUIZ, *Series históricas de capturas del atún rojo en las almadrabas del golfo de cádiz (siglos xvi-xxi)*, Collect. Vol. Sci. Pap. ICCAT, 67 (2012), pp. 139–174.
- [20] M. E. MANN and J. M. LEES, *Robust estimation of background noise and signal detection in climatic time series*, Clim. Ch., 33 (1996), pp. 409–445.
- [21] A. MATHIAS, F. GROND, R. GUARDANS, D. SEESE, M. CANELA, and H. DIEBNER, *Algorithms for Spectral Analysis of Irregularly Sampled Time Series*, J. of Stat. Soft., 11 (2004), pp. 1–27.
- [22] M. MUDELSEE, *TAUEST: A computer program for estimating persistence in unevenly spaced weather/climate time series.*, Computers & Geosciences, 28 (2002), pp. 69–72.
- [23] ———, *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*, Springer, (2010).
- [24] M. MUDELSEE, D. SCHOLZ, R. RÖTHLISBERGER, D. FLEITMANN, A. MANGINI and E. W. WOLFF, *Climate spectrum estimation in the presence of timescale errors*, Nonl. Proc. in Geophys., 16 (2009), pp. 43–56.
- [25] A. NAVA, *Fourier spectral analysis for unevenly spaced, average value, data*, Computers & Geosciences, 36 (2010), pp. 853–860.
- [26] A. NUTTALL, *Some windows with very good sidelobe behavior.*, IEEE Trans. on Acoust., Speech and Sig. Proces., 29 (1981), pp. 84–91.
- [27] E. PARDO-IGÚZQUIZA and F. RODRÍGUEZ-TOVAR, *Análisis espectral de series temporales de variables geológicas con muestreo irregular.*, Boletín Geológico y Minero, 124 (2013), pp. 323–337.
- [28] E. PARDO-IGÚZQUIZA and F. J. RODRÍGUEZ-TOVAR, *Implemented lomb-scargle periodogram: a valuable tool for improving cyclostratigraphic research on unevenly sampled deep-sea stratigraphic sequences.*, Geo-Marine Letters, 31 (2011), pp. 537–545.
- [29] E. PARDO-IGÚZQUIZA and F. J. RODRÍGUEZ-TOVAR, *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 (2012), pp. 207–216.
- [30] D. B. PERCIVAL and A. T. WALDEN, *Spectral analysis for physical applications*, Cambridge Univ Press, London, (1993).
- [31] J. POLANCO-MARTÍNEZ, *Aplicación de técnicas estadísticas en el estudio de fenómenos ambientales y ecosistémicos.*, PhD thesis, University of Basque Country, España., (2012).
- [32] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING and B. P. FLANNERY, *Numerical recipes in C*, Cambridge University Press, (1992).
- [33] M. B. PRIESTLEY, *Spectral analysis and time series.*, Academic Press Limited, London, (1981).
- [34] R DEVELOPMENT CORE TEAM, *R: A Language and Environment for Statistical Computing.*, R Foundation for Statistical Computing, Vienna, Austria, (2009).
- [35] D. H. ROBERTS, J. LEHAR, and J. W. DREHER, *Times series analysis with clean. I. derivation of a spectrum.*, The Astron. J., 93 (1987), pp. 968–989.
- [36] P. M. ROBINSON, *Estimation of a time series model from unequally spaced data.*, Stochastic Processes and their applications, 6 (1977), pp. 9–24.
- [37] E. SCANNELL JR and G. C. CARTER, *Confidence bounds for magnitude-squared coherence estimates.*, in Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'78., vol. 3, IEEE, (1978), pp. 670–673.
- [38] J. D. SCARGLE, *Studies in astronomical time series analysis. II- Statistical aspects of spectral analysis of unevenly spaced data.*, The Astrop. J., 263 (1982), pp. 835–853.
- [39] ———, *Studies in astronomical time series analysis. III- Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data.*, The Astrop. J., 343 (1989), pp. 874–887.
- [40] M. SCHULZ, *On the 1470-year pacing of dansgaard-oeschger warm events.*, Paleoceanography, 17 (2002), pp. 4–1.
- [41] M. SCHULZ AND M. MUDELSEE, *Manual de referencia REDFIT 3.5. Estimating red-noise spectra directly from unevenly spaced paleoclimatic time series.*, (2001).
- [42] ———, *REFIT: Estimating red-noise spectra directly from unevenly spaced paleoclimatic time series.*, Computers & Geosciences, 28 (2002), pp. 421–426.

- [43] M. SCHULZ and K. STATTEGGER, *SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series.*, Computers & Geosciences, 23 (1997), pp. 929–945.
- [44] A. SCHUSTER, *On the investigation of hidden periodicities with applications to a supposed 26 day period of meteorological phenomena.*, Terrest. Magn., 3 (1898), pp. 13–41.
- [45] H. P. A. VAN DONGEN, E. OLOFSEN, J. H. VAN HARTEVELT, and E. W. KRUYT, *A procedure of multiple period searching in unequally spaced time-series with the Lomb-Scargle method.*, Bio. Rhy. Res., 30 (1999), pp. 149–177.
- [46] P. VANICEK, *Further development and properties of the spectral analysis by least-squares.*, Astrop. Sp. Sc., 12 (1971), pp. 10–33.
- [47] H. VON STORCH and F. W. ZWIERS, *Statistical analysis in climate research.*, Cambridge University Press, Cambridge, U.K., (1999).
- [48] G. P. WEEDON, *Time-series analysis and cyclostratigraphy: examining stratigraphic records of environmental cycles.*, Cambridge Univ Press, Cambridge, (2003).
- [49] P. WELCH, *The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms.*, IEEE Trans. Audio & Electroac., 15 (1967), pp. 70–73.
- [50] Z. WU, N. E. HUANG, S. R. LONG, and C. K. PENG, *On the trend, detrending, and variability of nonlinear and nonstationary time series.*, Proc. National Academy Sciences, USA, 104 (2007), p. 14889.
- [51] P. YIOU, E. BAERT, and M. F. LOUTRE, *Spectral analysis of climate data.*, Surv. in Geophys., 17 (1996), pp. 619–663.

ANEXO A

A Estimación del espectro vía REDFIT

REFFIT viene acompañado de un excelente manual de nombre USAGE.PDF (véase /doc/ en redfit.zip, <http://www.ncdc.noaa.gov/paleo/softlib/redfit/redfit.html>), con instrucciones para instalación y uso, pero en lengua inglesa [41]. Por esta razón, se proporciona a continuación una breve descripción de cómo utilizarlo. El archivo de entrada lleva de nombre N1000.CFG (no es recomendable cambiarle de nombre debido a que REDFIT no funcionará). Su estructura básica es la siguiente:

```
&cfg
  fnin = 'nombre del archivo de entrada',
  fnout = 'nombre del archivo de salida',
  nsim = ,
  mctest = ,
  ofac = ,
  n50 = ,
  iwin =
/

!iwin =
!      0: Rectangular
!      1: Welch
!      2: Hanning
!      3: Triangular
!      4: Blackman-Harris
```

Los primeros dos parámetros, *fnin*, *fnout*, contienen los nombres de los archivos de entrada (que incluyen la serie temporal que se va a estudiar, la cual debería estar en texto plano, delimitados por espacios o por *tabs* y contener a la vez un sólo tiempo y dato por línea. Para mayor información véase sección 2.2 del manual USAGE.PDF de REDFIT [41]) y de salida (acepta cualquier nombre), respectivamente. Es importante que ambos archivos están entre comillas simples ' '. Del siguiente par de parámetros, *nsim* y *mctest*, el primero se refiere al número de simulaciones de Monte-Carlo (se recomiendan al menos 1000 simulaciones); el segundo, se utiliza T para activar el test de Monte-Carlo (use F, para desactivarlo), el cual estima los niveles de significación estadística de los picos espectrales [41, 42]. El parámetro *ofac* controla el número de frecuencias utilizadas en la estimación del periodograma de Lomb-Scargle; por lo general, se usa un valor entre 2.0 y 4.0; sin embargo, tenga en cuenta que, a mayor valor, mayor será el tiempo computacional en la estimación del espectro; finalmente, *n50* indica el número de segmentos WOSA y *iwin* determina el tipo de ventana (*taper*) [41].

Después de ejecutar REDFIT exitosamente (en caso contrario se genera un archivo de nombre *redfit.log*, que

muestra los errores y otras advertencias), se obtienen dos tipos de salidas. La primera en pantalla contiene lo siguiente:

```
N =
t(1) =
t(N) =
<dt> =
Nout =

Required RAM =

***** tau *****\n

the dof
```

donde *N* es el número de elementos de la serie temporal a estudio, *t(1)* y *t(N)* son el primer y último valor de los tiempos de la serie, *< dt >* es la media aritmética de las diferencias de los tiempos, *Required RAM* es la cantidad de memoria RAM necesaria para estimar el espectro, *tau* es el coeficiente de persistencia de la serie temporal bajo análisis [22, 42], y finalmente, *the dof* (*the degree of freedom*) son los grados de libertad del espectro estimado. Un espectro crudo tiene 2 grados de libertad, pero uno suavizado tiene un dof mayor que 2. Algunos autores sugieren que al suavizar el espectro se obtengan mucho más que 2 grados de libertad (por ejemplo, 6 u 8), con el objeto de eliminar posibles falsos picos espectrales [23, 48]. Sin embargo, hay que tener cuidado, porque un espectro demasiado suavizado podría eliminar picos espectrales que pudieran representar señales reales.

La segunda salida es un archivo en texto plano, cuyo nombre viene dado por el parámetro *fnout*. Este fichero contiene un encabezado de unas 50 líneas con comentarios (# indica comentario) referentes a los parámetros de configuración, valores iniciales y el significado de las variables (entre 10 y 14 columnas, lo cuál depende si se activa el test de Monte-Carlo para determinar la significación estadística de los picos espectrales). Entre estas variables, destacan las frecuencias en las que se calculan el PLS (columna 1), el espectro corregido (columna 3) y suavizado (en caso de haber definido un *n50* mayor que uno o aplicado alguna ventana de suavizado), el espectro teórico de un proceso AR1 con un nivel de significación del 0.95 (columna 9) y el nivel de significación de 0.95 usando Monte-Carlo (columna 13) –en caso de que se haya activado *mctest = T*–.

Antes de utilizar REDFIT, es recomendable tener en cuenta varias pautas. La primera se refiere al número de elementos de la serie temporal que se va a estudiar, *N*; en principio, no existe límite en el número de elementos; éste viene dado, más bien, por la memoria del ordenador utilizado. La segunda pauta –y es muy importante tenerla en cuenta– es que los tiempos de muestreo de la serie temporal que se estudia deben de seguir una secuencia monótona estrictamente creciente (*i.e.*, $t_1 < t_2 < \dots < t_N$, donde t_1 es el

primer valor del tiempo en el archivo de entrada mientras que t_N es el último valor) [41, 42]. La tercera pauta es que aunque REDFIT fue programado para utilizarse en Windows, es posible compilar el código fuente para trabajar en otros sistemas operativos⁵. La cuarta, es que los auto-

res de REDFIT recomiendan utilizar el software Gnuplot (<http://www.manfredmudelsee.com/soft/redfit/>) para graficar la salida de REDFIT. Sin embargo, en este artículo presentamos un sencillo programa en lenguaje R [34] (véase anexo B) para visualizar la salida de REDFIT.

ANEXO B

B Programa para graficar la salida de REDFIT

```
#####
# Este script grafica la salida de REDFIT [Schulz & Mudelsee 2002]
#####
#:: Josué M. Polanco-Martínez (josue.m.polanco@gmail.com)
#:: Código liberado bajo la Licencia Pública General de GNU (GPL)
#:: http://www.gnu.org/licenses/gpl-3.0.html
#:: Bilbao, Agosto 2014.
#####
#:: Esta función transforma a decibelios
source("./decibel.R")
#####

#:: Se lee la salida n1000.red de REDFIT y se definen variables.
#####
redfitdat <- read.table("./n1000.red")
len = length(redfitdat[,1])

vars <- 1:dim(redfitdat)[2]
names(vars) <- c("Freq", "Gxx", "Gxx_cor", "Gred_th",
                "<Gred>", "CorrFac", "80%-Chi2",
                "90%-Chi2", "95%-Chi2", "99%-Chi2",
                "80%-MC", "90%-MC", "95%-MC",
                "99%-MC")
#####
#:: Se grafica y guarda
#####

pdf_file <- ("grafica_salida_redfit.pdf")

ranche <- range(db(redfitdat[,3]))
Y1 <- ranche[1] - 40
Y2 <- ranche[2] + 20
pdf(file=pdf_file, height=4, width=5)
plot(redfitdat[,1], db(redfitdat[,3]), t="l",
      xpd=FALSE, xlab="", ylab="", ylim=c(Y1, Y2))

title(main="", xlab="Frecuencia [ciclos por A\`o]",
      ylab="Amplitud espectral [dB]")
points(redfitdat[,1], db(redfitdat[,10]), t="l", lty=2, col="red")
points(redfitdat[,1], db(redfitdat[,14]), t="l", lty=3, col="blue")
```

⁵El autor de este artículo compiló el código fuente de REDFIT para trabajar en Linux; si tiene interés en cómo llevar a cabo esta tarea, póngase en contacto vía correo-e.

```
#####
#::: Drawing the significant spectral peaks (above 99% CL)
the.points <- which(redfitdat[,3] >= redfitdat[,14])
cat("the points", the.points)
len.points <- length(the.points)
freq.points <- array(0, len.points)
psdDC.points <- array(0, len.points)
freqs <- redfitdat[,1]
psdDC <- redfitdat[,3]

for (i in 1:len.points) {
  freq.points[i] <- freqs[the.points[i]]
  psdDC.points[i] <- psdDC[the.points[i]]
}

points(freq.points, db(psdDC.points), pch=19, cex=0.5)
text(freq.points, db(psdDC.points)+.5, labels=c(1:len.points),
      pos=4, cex=.5)

legend("bottomleft", paste(1:len.points, "->",
  round(1/freq.points[1:length(freq.points)]), digits=1),
  "A\'os"), bty="n", cex=.75)

legend("topright", lty=c(2,3), c("99% NF v\'a AR1", "99% NF v\'a MC"),
  col=c("red", "blue"), bty="n", cex=0.7)

dev.off()
```