

TRF SOBRE DATOS POSTMUESTRADOS

V. M. Peñafiel¹

Carrera de Física, FCPN (UMSA)
La Paz, Bolivia

RESUMEN

Se exhibe y revisa tres principales contaminantes del espectro de potencias correspondiente a datos experimentales cuyos registros fueron obtenidos previamente y sin el propósito específico de ser analizados via la transformada digital de Fourier. Para este tipo de registros, algunas técnicas analíticas del procesamiento de señales digitales (PSD) están todavía disponibles; sin embargo, los principales filtros contra el ruido y la tergiversación (“aliasing”) ya no pueden ser aplicados. Se hace énfasis, por tanto, en el tratamiento de tales problemas, especialmente en el de la detección de frecuencias espurias (alias), imposible para el PSD cuando se desconoce si la señal digital ha sido correctamente muestreada. Se prueba que el empleo de un método ajeno a las operaciones relacionadas con la transformada rápida de Fourier (TRF), como auxiliar analítico, permite descubrir las tergiversaciones conspicuas.

1. INTRODUCCIÓN

Las sofisticadas y poderosas técnicas del Procesamiento de Señales Digitales están principalmente orientadas a producir registros limpios de ruidos y efectos contaminantes aún antes de un análisis de señal propiamente dicho [1] [2] [3]. El empleo extendido de filtros a nivel de muestreo, traducidos directamente en circuitos electrónicos digitalizadores, veloces y eficientes, llevan a la práctica el “prevenir antes que lamentar” en materia de análisis armónico; y las aplicaciones están ampliamente extendidas a nivel comercial, industrial y científico en gran diversidad de áreas particulares [1]. También en Física, por supuesto, el empleo de analizadores digitales en la obtención de datos apropiados para el estudio de fenómenos periódicos armónicos o anarmónicos es común y hasta rutinario en ciertos campos.

Desafortunadamente, un buen conjunto de problemas relativos al tratamiento de datos experimentales tiene que ver con registros cuyos procedimientos de medición no estaban diseñados para el PSD, no eran apropiados para ello, no constituyen digitalización de fenómenos analógicos o simplemente provienen de procedimientos de medición parcial o totalmente desconocidos. En general, tales registros se presentan al análisis solamente para obtener información adicional a cerca de periodicidades posibles o esperadas o a cerca de tendencias espectrales con implicaciones posteriores.

Por tanto, estas series temporales o espaciales se hallan impregnadas de ruidos indeterminados, frecuencias espurias y otros contaminantes para los cuales el PSD puede no rendir una clara resolución. De todos modos, un análisis de datos postmuestreados debe ser

capaz de reducir significativamente el efecto de esas interferencias antes de la extracción de conclusiones. En este trabajo se considera tres de los efectos más recurrentes: el ruido aleatorio (blanco) el ruido $1/f$ (rosado) y la tergiversación (*aliasing*) de señales de alta frecuencia (Fig. 1.).

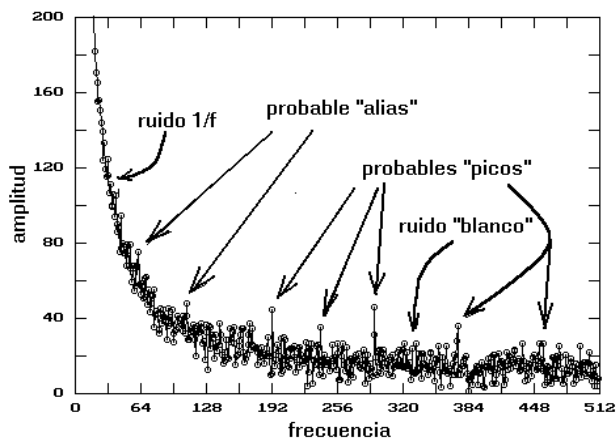


Figura 1. Espectro de amplitudes correspondiente a un típico registro de datos postmuestreados.

En una serie de datos postmuestreados, (i) el nivel de ruido y (ii) el intervalo de muestreo (Δt ó Δx) son características dadas e irrevisables, de modo que la sugerencia de “emplear intervalos de muestreo y filtros más adecuados” no es aceptable. El análisis se orienta a obtener la mayor información posible con los datos caracterizados de esa manera o decidir si el registro es francamente inservible para propósitos de tratamiento de señal. Por otra parte, aunque usualmente el procesamiento de datos desde el punto de vista del análisis armónico es por sí mismo un tema amplio

¹Email: vmiguel@fiumsa.edu.bo

y lleno de sutiles detalles, que requiere -a veces- el desarrollo de métodos específicos a un tipo particular de registro, en términos generales lo que se busca es la *detección* de determinadas periodicidades ("picos") y, eventualmente, la medición de de sus frecuencias dentro del margen de error permitido por los datos; menos frecuente y necesaria es la determinación de amplitudes y diferencias de fase. Otras propiedades que el espectro de potencias o de amplitudes puede revelar, como la presencia de fenómenos caóticos, son menos susceptibles al tratamiento cuantitativo, por lo cual no ocupará nuestra atención.

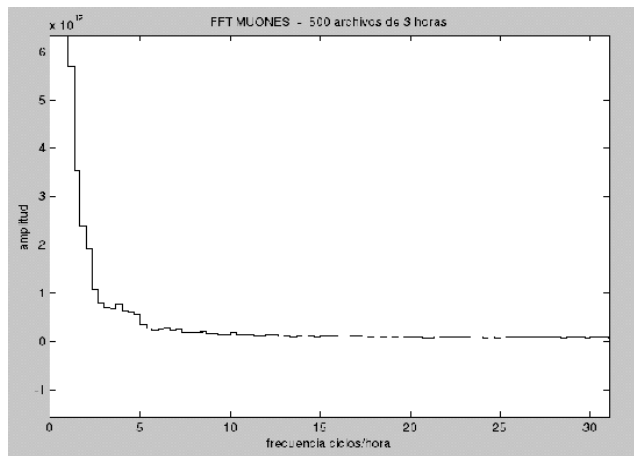


Figura 2. Primeras 32 amplitudes de un espectro para 10800 eventos de rayos cósmicos por hora (con ruido rosado).

2. RUIDOS BLANCOS Y ROSADOS

Los más perjudiciales contaminantes en un registro experimental son las fluctuaciones aleatorias con energía distribuida sobre todo el espectro, el ruido blanco", y otro, cuyo origen no está muy claramente explicado, cuya amplitud aumenta en las bajas frecuencias según $1/f$ y que se conoce como ruido rosado"[1].

De este último hay muy poco que decir cuando es inherente al proceso de medición o al fenómeno estudiado; en ocasiones, sin embargo, cierto tipo de ruido rosado se filtra en la rutina de cómputo cuando ésta usa iterativamente ciertas posiciones de memoria -principalmente para economizarla- sin una previa limpieza. Cuando los registros de datos son muy grandes, es muy recomendable calibrar el programa de cálculo con una base de datos iguales a una constante y del mismo tamaño que el registro experimental para verificar su correcto comportamiento a bajas frecuencias y, claro, es más recomendable aún el uso de rutinas con código fuente conocido. Por ejemplo, el espectro de amplitudes de la figura 2., obtenido mediante un programa cerrado"[6], muestra un claro decaimiento del tipo $1/f^n$; sin embargo, el mismo registro de 10800 datos procesado con una rutina calibrada revela que tal decaimiento es inexistente (Fig. 3.). El valor elevado

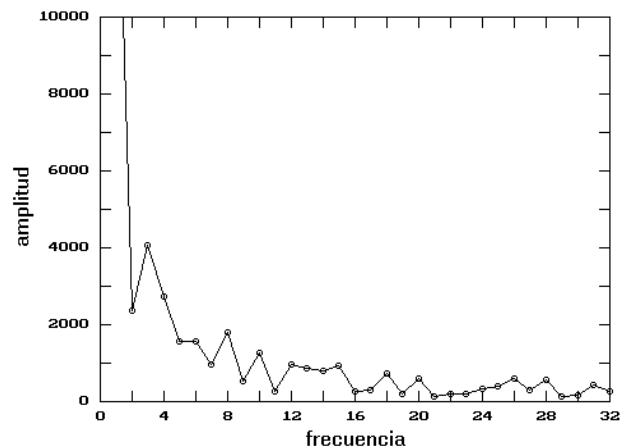


Figura 3. Espectro de amplitudes correspondiente al mismo registro empleado para la Fig. 2. (sin ruido rosado).

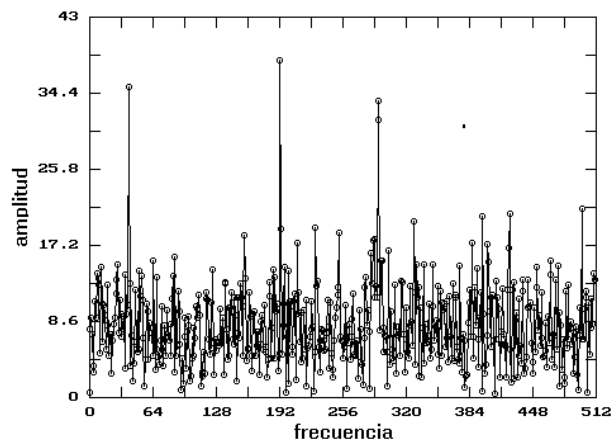


Figura 4. Espectro de amplitudes correspondiente al mismo registro empleado para la Fig. 1. (sin ruido rosado).

en el origen puede reducirse a uno pequeño o cero restando una constante (v.g., el promedio del registro) a la base de datos original. El mismo procedimiento, aplicado a los datos usados para obtener el espectro de la figura 1., rinde el de la figura 4., donde parte del ruido contaminante aledaño a los "picos" interesantes ha desaparecido también.

En muchos casos, existe ruido desigualmente distribuido sobre el intervalo de frecuencias; o ruido con marcadas fluctuaciones de amplitud, con máximos que pueden llegar a ser confundidos con picos de señal. No todas las interferencias aleatorias constituyen, pues, ruido blanco". En cualquier caso, todo tipo de ruido es perjudicial para el análisis cuando sus niveles de amplitud (o potencia) son lo suficientemente altos como para eclipsar información útil.

En efecto, aunque en principio es posible reducir el efecto del ruido aleatorio, como es usual en el PSD, fraccionando el registro total -si es bastante grande- en segmentos más cortos, efectuar sobre ellos la TRF y luego

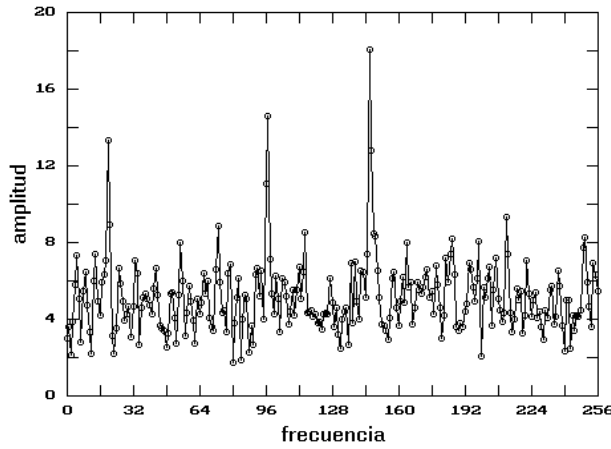


Figura 5. Fraccionamiento en 3 segmentos filtrados para los datos correspondientes al espectro de la Fig. 4.

promediar los espectros resultantes, la limitación a tal procedimiento la da el teorema de Parseval:

$$\sum_{i=0}^{n-1} y_i^2 = \frac{1}{n} \sum_{k=0}^{n-1} |F_k|^2 \quad (1)$$

(la potencia de la serie de datos se distribuye sobre todas las frecuencias admitidas, independientemente de si constituyen señal o ruido). El fraccionamiento de la serie debilita los picos de señal porque la potencia total también se fracciona. Así, mientras el espectro de la Fig. 3., cuando su serie de datos ha sido fraccionada en tres segmentos, muestra una reducción aceptable en las amplitudes ruidosas (y permite eliminar posibles picos al no mantener éstos su posición y tamaño relativos en el espectro) (Fig. 5.), la misma serie, fraccionada en 12 partes, muestra ya un deterioro notable debido a la reducción en las amplitudes de los picos de señal (Fig. 6.).

En consecuencia, se puede lograr una efectiva reducción de ruido aleatorio en aquellas afortunadas ocasiones en las que se dispone de varios registros *independientes* $\{x_i, y_i\}_n$ obtenidos con el mismo procedimiento de medición. El promedio de los espectros de amplitud $\{\langle F_k \rangle\}_n$ dados por la TRF de esos registros proporciona, entonces, una versión más "limpia" del espectro que contiene a los picos de señal interesantes porque la superposición afecta, ahora, a las amplitudes no coincidentes (aleatorias) pero no disminuye la potencia contenida en las componentes periódicas (Fig. 7.).

3. EL PROBLEMA DE LA TERGIVERSACIÓN DE SEÑAL

Si el propósito principal del tratamiento de los datos es la búsqueda de periodicidades significativas, entonces la tergiversación constituye el problema central a resolver pues, en principio, cualquier pico de señal en el espectro para datos postmuestreados es sospechoso de

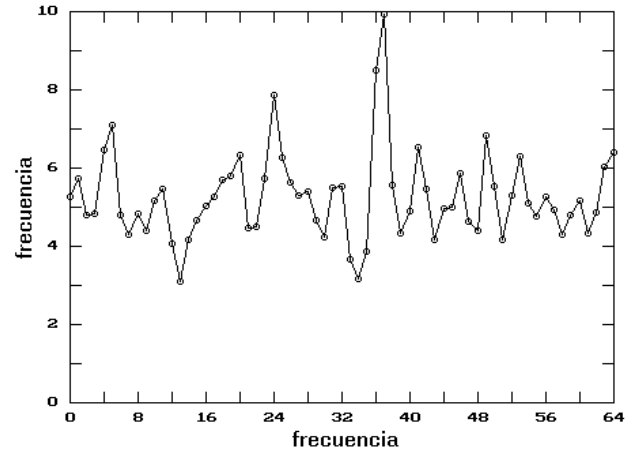


Figura 6. Fraccionamiento en 12 segmentos filtrados (de 128 puntos cada uno) para los datos correspondientes al espectro de la Fig. 4.

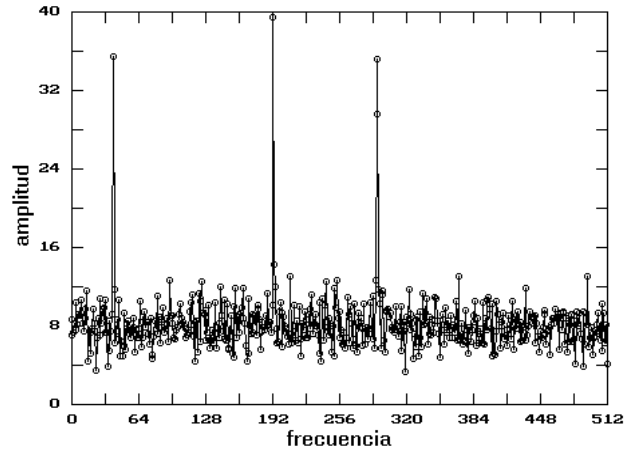


Figura 7. Superposición (promedio) de las TRF para 6 registros independientes sobre el mismo procedimiento experimental de la Fig. 4.

ser un "alias".

Los alias son picos espurios que corresponden a frecuencias altas, digamos ν_a , y que son reflejados sobre una frecuencia baja

$$\nu_b = 2n\nu_0 - \nu_a \quad (2)$$

donde ν_0 es la frecuencia de Nyquist y n algún número entero. La figura 8. ilustra claramente el efecto: una señal de alta frecuencia (período T) es digitalizada con un lapso de muestreo $\Delta t > T$; el resultado es una serie temporal (puntos oscuros) que representa a una señal de baja frecuencia; esta señal -no la original- es la que realmente interviene en el mecanismo de la transformada digital y conduce a un pico falso en el espectro respectivo (Fig. 9.).

Evidentemente, pues, en los hechos la señal de entrada es **sustituida** por otra de menor frecuencia. En el dominio frecuencial del análisis de Fourier, esa

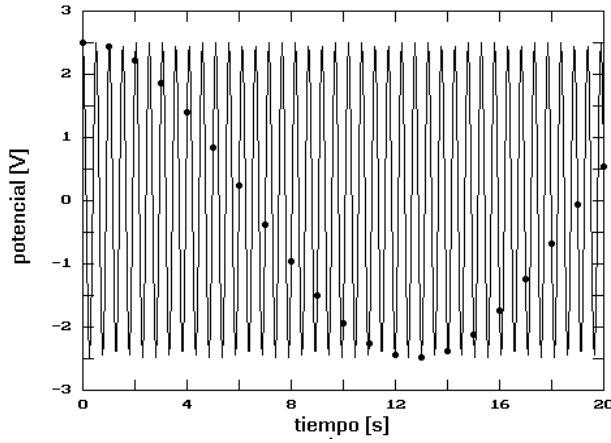


Figura 8. Señal de período $T=0.53$ [s] mostrada cada segundo. La señal digitalizada es la que se corresponde con la serie de círculos llenos (muy obviamente de frecuencia más baja).

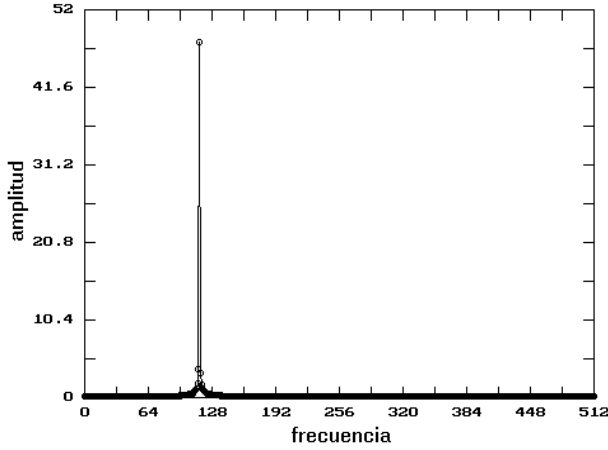


Figura 9. Espectro de amplitud para la señal digitalizada de la Fig. 8. El pico en $\nu = 116$ es un "alias" del que debería aparecer en $\nu = 1932$ para un registro con 1024 datos ($2048-1932=116$).

señal de entrada es invisible. Un inconveniente adicional de la tergiversación, cuando el proceso de digitalización no ha sido apropiado o se desconoce, es que todo pico de señal es, en principio, sospechoso de ser alias de alguna otra señal "verdadera". El PSD no permite inferir la frecuencia de la señal verdadera pues la ec. (2) admite infinitas soluciones para ν_b y ν_0 dadas.

Existe, sin embargo, un muy antiguo método estadístico de estimación espectral basado sobre la propiedad de interferencia de dos o más trenes de onda y que, por brevedad, lo denominaremos "método Maxvar". El algoritmo asociado es el siguiente [5] [6]:

Suponiendo que los datos tienen la forma general

$$y_i = \sum_{r=1}^k A_r \text{sen}(\omega_r i - \phi_r) \quad (3)$$

con $\omega_r = 2\pi/T_r$ y $\Delta t = 1$ (unidades arbitrarias), los promedios sobre términos como $z_r(i) = A_r \text{sen}(\omega_r i - \phi_r)$ son, estadísticamente,

$$\langle z_r \rangle \propto \int_0^T \text{sen}(\omega_r x - \phi_r) dx = 0$$

cuando las $z_r(i)$ están fuera de fase o, más bien, en las diferentes fases de la función trigonométrica y

$$\langle z_r \rangle = \langle A_r \text{sen}(\theta_r) \rangle = A_r \text{sen}(\theta_r)$$

con $\theta_r = c - \phi_r$ si las cantidades $z_r(i)$ están en fase, esto es, si sus argumentos se relacionan mediante alguna constante c según $\omega_r i \simeq 2\pi + c$.

Esto implica que cuando los datos son ordenados en arreglos de β filas y α columnas y promediados por columnas:

$$\begin{array}{cccc} y_1 & y_2 & \dots & y_\alpha \\ y_{\alpha+1} & y_{\alpha+2} & \dots & y_{2\alpha} \\ y_{2\alpha+1} & y_{2\alpha+2} & \dots & y_{3\alpha} \\ \dots & \dots & \dots & \dots \\ y_{(\beta-1)\alpha+1} & y_{(\beta-1)\alpha+2} & \dots & y_{\beta\alpha} \end{array}$$

$$m_1^\alpha \quad m_2^\alpha \quad \dots \quad m_\alpha^\alpha$$

las medias m_u^α son pequeñas y homogéneas entre sí (cuando los términos de las columnas están fuera de fase) o están dispersas en las diferentes fases de una señal ondulatoria (cuando los términos de cada columna están en fase). Consecuentemente, si se calcula la varianza

$$s_\alpha^2 = \langle (m^\alpha - m_\alpha)^2 \rangle \quad (4)$$

con $m_\alpha \equiv \langle m^\alpha \rangle = \sum_{u=0}^{\alpha} m_u^\alpha$, y la serie de razones

$$\lambda_\alpha = \frac{s_\alpha}{s_y} \quad (5)$$

donde s_y es la desviación típica del registro entero, λ_α debe ser un máximo relativo cuando α marca un período o un múltiplo entero del período T_r en el arreglo con α columnas. Para una señal armónica, es fácil demostrar que $s_\alpha \approx A_\alpha/\sqrt{2}$ y, por tanto, para datos de la forma (3), se debe cumplir la restricción aproximada

$$s_y^2 \approx \sum_{r=1}^k \frac{A_r^2}{2} \Rightarrow \sum_{r=1}^k \lambda_r^2 \approx 1 \quad (6)$$

para las λ que constituyen máximos relativos.

El análisis Maxbar es lento y tedioso, en general, por lo que requiere de ágiles rutinas de cálculo y graficado. Por ejemplo, el diagrama Maxvar para la señal tergiversada de la figura 8., que se muestra en la figura 10., exhibe repetitivamente todos los picos compatibles con la única periodicidad contenida en el registro. Las figuras 11. y 12. contienen el detalle

de las alturas de las λ_α que son relevantes al análisis. No obstante que existen máximos relativos compatibles con la frecuencia $\nu = 116$ proporcionada por la TRF, el hecho de que las alturas de las λ_α para $\alpha = 53, 106, 159$ etc., sean iguales a la unidad es una clara y definitiva indicación de que la frecuencia $\nu = 116$ (correspondiente a un período $T = 8,83$) es un alias. La frecuencia correcta se halla haciendo que la ec. (2) se satisfaga para el período menor a la unidad del cual $T=53$ es múltiplo, esto es, como $1024/53 \approx 1932$ y $2(1024) - 1932 = 116$, deduciríamos que que la frecuencia correcta presente es 1932 y no 116. En muchos casos, el procedimiento es ya útil cuando permite descartar picos del espectro detectados como alias, aún si las frecuencias "verdaderas" no fueran unívocamente ubicables.

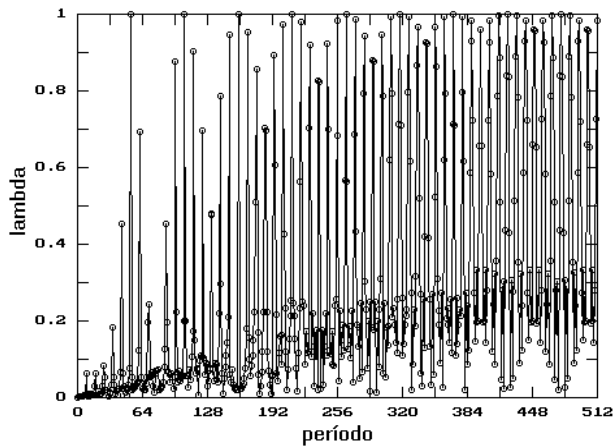


Figura 10. Diagrama Maxvar para la señal digitalizada de la Fig. 8. El pico en $T = 53$, por tener una altura $\lambda = 1$ indica que la única frecuencia presente debe ser compatible con algún múltiplo de ese período.

Así, para los datos considerados en este artículo, cuya TRF, menos afectada de ruidos, se muestra en la figura 7., el procedimiento Maxvar rinde el nutrido diagrama de la figura 13. En la figura 14. aparece el detalle de los primeros 60 puntos. Los picos en $\nu = 193$ y $\nu = 293$ del espectro TRF (Fig. 7) quedan confirmados con los picos en $T = 7, 14, 21, \dots$ (múltiplos de $T = 3,5 \approx 1024/293$) y $T = 53$ (múltiplo de $T = 5,3 \approx 1024/193$), pero el pico en $\nu = 40$ de la TRF debería aparecer bastante bien representado en el Maxvar por ejemplo en $T = 256$ (múltiplo de $T = 25,6 \approx 1024/40$); sin embargo, la figura 15. no muestra un máximo relativo para ese período. La conclusión es que la frecuencia $\nu = 40$ del espectro TRF es un alias.

No es plausible que el método Maxvar pueda ser usado sistemáticamente para un análisis espectral efectivo. Sin embargo, los ejemplos presentados lo hacen un excelente *auxiliar* en la inspección detallada de periodicidades significativas dentro de registros experimentales no necesariamente orientados hacia el procesamiento de señales. Su efectividad, naturalmente,

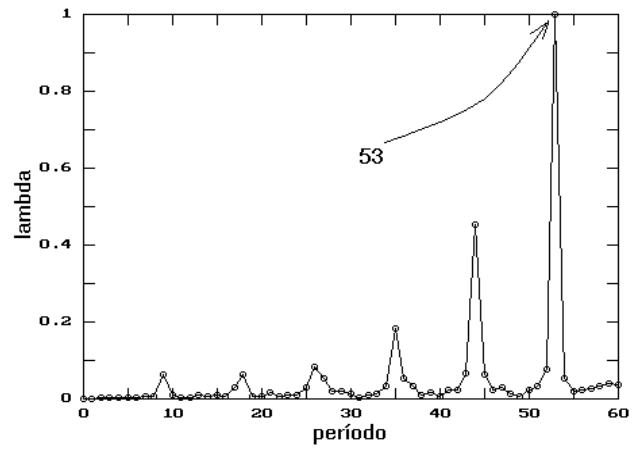


Figura 11. Detalle del diagrama Maxvar de la Fig. 10. mostrando los primeros 60 puntos el pico en $T = 53$ es ahora evidente.

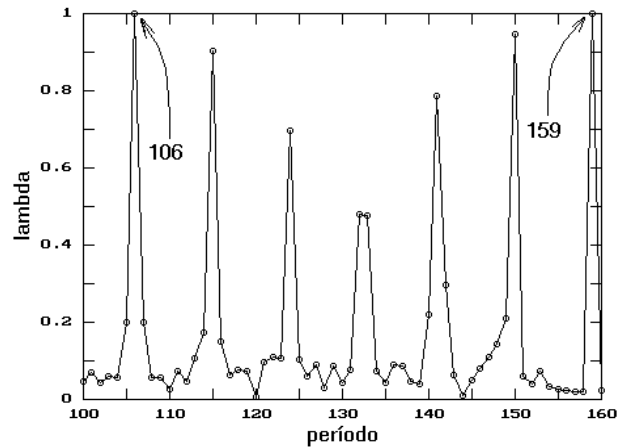


Figura 12. Detalle del diagrama Maxvar de la Fig. 10. con la confirmación de que el único período presente es un submúltiplo de $T = 53$.

está condicionada también a la calidad precisión y tamaño de las series de datos disponibles.

El análisis puede, eventualmente, ser mejorado también con el cauteloso empleo de filtros postmuestres. En buena parte de los casos, principalmente porque el contenido de los datos que nos ocupan es casi siempre desconocido (en términos de propiedades de señal) un filtro de los tipos Hanning, Hamming o Blackman

$$\begin{aligned} w(i) &= \frac{1}{2} [1 - \cos(\frac{2\pi i}{n})] \\ w(i) &= 0,54 - 0,46\cos(\frac{2\pi i}{n}) \\ w(i) &= 0,42 - 0,5\cos(\frac{2\pi i}{n}) + 0,08\cos(\frac{2\pi i}{n}) \end{aligned} \quad (7)$$

($0 \leq i \leq n$) [1] [2] [3] debe ser suficiente cuando ciertos picos se encuentran muy próximos o se sospecha que algunas amplitudes de ruido se deben a filtraciones producidas por efecto de los bordes (tamaño) del registro. Los datos filtrados $y(i) = x(i)w(i)$ admiten todo el análisis precedente y la elección de uno de los

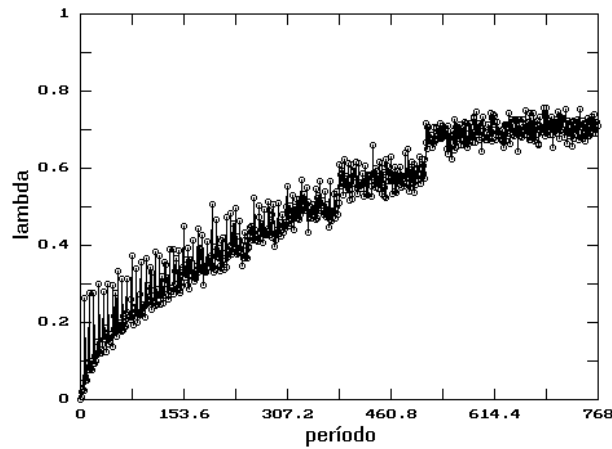


Figura 13. Diagrama Maxvar para el mismo registro usado en las TRF de la Fig. 3.

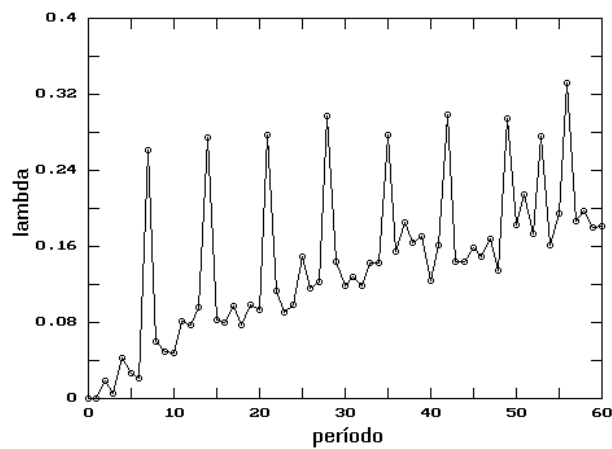


Figura 14. Detalle del diagrama Maxvar de la Fig. 13. Los picos sucesivos en 7, 21, 28, ... y el pico en 53 concuerdan con las frecuencias 293 y 193 de la TRF, respectivamente.

filtros (7) depende enteramente de la bondad de los resultados que se obtengan (ver Figs. 5. y 6.).

4. CONCLUSIONES

Muchos más ejemplos de los que se ha presentado podrían ilustrar con mayor detalle los particulares problemas que se presentan en un análisis espectral sobre datos "crudos" de los cuales se desea extraer información sobre periodicidades significativas. Una gran parte de tales registros postmuestreados han sido obtenidos sin gran preocupación por su posible tratamiento de señal y, por tanto, su digitalización no ha implicado técnicas de filtrado de altas frecuencias, ruido aleatorio o ruido rosado.

Las técnicas asociadas al análisis espectral de Fourier, poderosas y precisas cuando los datos son preparados con ese propósito, pueden rendir resultados imprecisos o ambiguos cuando el muestreo de las mediciones se desconoce. En esos casos, la pertinencia de los equipos de medición ya no es un tema de discusión y el

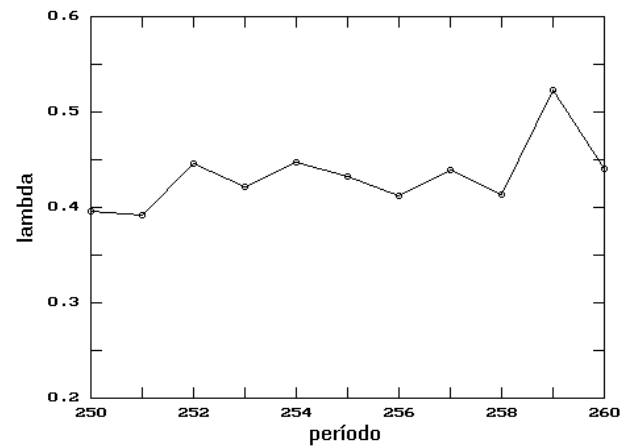


Figura 15. Detalle del Maxvar de la Fig. 13. No existe un máximo relativo en 256 (esperado de acuerdo con la TRF).

análisis debe basarse enteramente sobre procedimientos numéricos apropiados. Entre las técnicas conocidas, generalmente las de atenuación de ruido aleatorio y el empleo de rutinas de cómputo debidamente probadas, incluyendo las de filtrado de señal (ec. (7)) todavía pueden ser usadas con mucho éxito en combinación con la transformada rápida de Fourier.

Sin embargo, el efecto de tergiversación de señal, verdadera o potencialmente implicada en un registro de datos, plantea un problema insoluble para la transformada de Fourier digital; sobre todo porque ésta actúa sobre el dominio actual de frecuencias de la señal digitalizada de entrada; todas las recomendaciones en relación a evitar la presencia de señales con frecuencias superiores a la frecuencia de Nyquist están orientadas precisamente a prevenir las ambigüedades consecuentes. Por tanto, una vez que el muestreo ha tenido lugar, es poco lo que el análisis espectral basado sobre la TRF puede hacer para detectar frecuencias espurias.

Se ha tratado de mostrar, a lo largo de este trabajo, que la combinación de las técnicas de la TRF con el método Maxvar de estimación espectral (ya obsoleto para ese fin) puede lograr, en principio, la detección de periodicidades tergiversadas (alias) en el espectro asociado a registros postmuestreados, permitiendo el aislamiento de periodicidades significativas para un análisis posterior. La importancia de diseñar mecanismos de detección de alias radica, justamente, en garantizar la exactitud de las conclusiones que se desprenden luego del tratamiento espectral, que pueden tener implicaciones fatales al momento de interpretar físicamente el contenido de los registros experimentales que contienen periodicidades interesantes.

El método Maxvar, por su naturaleza estadística y por actuar solamente en el dominio de los períodos de las señales, es ajeno a las propiedades de la TRF y procedimientos asociados. En muchos casos, es muy sensible

inclusive a periodicidades con nivel señal-ruido muy bajo. Además, su programación para el cómputo electrónico es muy simple, como se sigue de la tabulación de datos y las ecs. (3) á (6). El número de posibles arreglos para n datos es, claro, n (de los cuales sólo $n/2$ proporcionan amplitudes significativas) por tanto, las rutinas de cálculo son muy rápidas para registros de tamaño razonable. Y, finalmente, puesto que ha de ser empleado únicamente como un auxiliar en el análisis global, los tediosos procedimientos que su empleo como método de estimación implican, ya no son necesarios.

REFERENCIAS

- [1] Smith S. W., *"The Scientist and Engineer's Guide to Digital Signal Processing"* Cal. Tech. Publishing, San Diego CA. (1999).
- [2] Brigham O. E., *"The Fast Fourier Transform"* Prentice Hall, Englewood (1974).
- [3] Cerna M. and Harvey A. E., *"The Fundamentals of FFT-Based Signal Analysis and Measurement"* Application Note 041, National Instruments Co., 2000.
- [4] Whittaker E. T. and Robinson G., *"Calculus of Observations"* Blackie & Son, London (1924).
- [5] Peñafiel V. M., *"Introducción al Tratamiento de Datos Experimentales"* UMSA, la Paz (1997).
- [6] Velarde A., Comunicación privada.