

PRUEBAS DE ALINEALIDAD POR DATOS SUBROGADOS SOBRE SERIES EXPERIMENTALES

NONLINEAR TESTS BY SURROGATE DATA ON EXPERIMENTAL SERIES

V. M. PEÑAFIEL[†]

Instituto de Investigaciones Físicas, Carrera de Física
Universidad Mayor de San Andrés
c. 27 Cota-Cota, Campus Universitario, Casilla de Correos 8639
La Paz – Bolivia

(Recibido 13 de Junio de 2012; aceptado 23 de Octubre de 2012)

Se introduce el método de datos subrogados para el ensayo de alinealidad sobre series temporales experimentales, en principio, como una etapa previa a ensayos de comportamiento caótico. Se adaptó un conjunto de algoritmos computacionales orientados específicamente a ese propósito y se los aplicó a registros geomagnéticos (componentes D, H y Z; tres series de 2048 datos, tomados en lapsos de 1 minuto) para verificar su capacidad de discriminación en el proceso de aislar series con características interesantes desde el punto de vista de la dinámica no-lineal. Se aplica los métodos de Fourier y de temple simulado para la construcción de registros subrogados. En este último caso, restringidos por la condición de invariancia de la función de autocorrelación, el muestreo de los coeficientes de predicción alineal exhibe un poder de discriminación muy notable al ser comparado con los de las otras pruebas.

Descriptores: Análisis de series temporales — dinámica no-lineal y caos

Código(s) PACS: 05.45.Tp, 05.45.-a

ABSTRACT

The method of surrogate data is introduced to test the nonlinear character of experimental time series, in particular, as a previous stage to test chaotic behaviour. We used several algorithms specifically oriented towards this purpose and applied to geomagnetic records (components D, H, Z and 3 time series with 2048 points each and taken at intervals of 1 minute) in order to test their discrimination capacity while studying series that could be interesting (because of their nonlinear character). We apply Fourier and simulated annealing methods to construct surrogate data. For the latter, and restricted by the invariance of the autocorrelation function, the sampling of the coefficients of the nonlinear prediction shows a remarkable discrimination capacity as compared to other methods.

Subject headings: Time series analysis — nonlinear dynamics and chaos

1. INTRODUCCIÓN

Los ensayos de alinealidad no son muy frecuentes en Física porque las oscilaciones erráticas, características de las series temporales que provienen de procesos de medición, son consideradas, simplemente, como ruido estocástico. Cuando existe sospecha de estructuras complejas, como determinismo caótico, el cálculo de coeficientes de Lyapunov y análisis relacionados sirve, de paso, como un ensayo de linealidad (o alinealidad).

Desde la perspectiva estricta del tratamiento de datos, las técnicas propias de dinámica no lineal tienden a mostrar resultados inciertos cuando las series de datos son “ruidosos”, esto es, cuando la amplitud de las oscilaciones aleatorias es comparable o sobrepasa a la de los efectos no lineales, por lo cual

parece importante respaldarlos con técnicas de ensayo alternativas.

Pero, además, a lo largo de este trabajo se muestra que el método por datos subrogados puede facilitar la selección de registros con propiedades interesantes para análisis posteriores.

En efecto, las figuras 1(a), 1(b) y 1(c) grafican el registro simultáneo de las componentes D, H y Z del campo geomagnético en superficie (Ricaldi 2012) (2048 lecturas en lapsos de 1 minuto). Se trata de cortes convenientes desde registros más largos, evitando efectos “de borde” en sus extremos y en un número que facilite la aplicación repetida de la transformada rápida de Fourier (TRF). Se aprecia las tendencias macroscópicas con pequeñas oscilaciones erráticas superpuestas. El objetivo del análisis subsiguiente es el de detectar -en tales series- efectos alineales, si los hubiera. La aplicación del análisis

[†]vmiguel@fiums.edu.bo

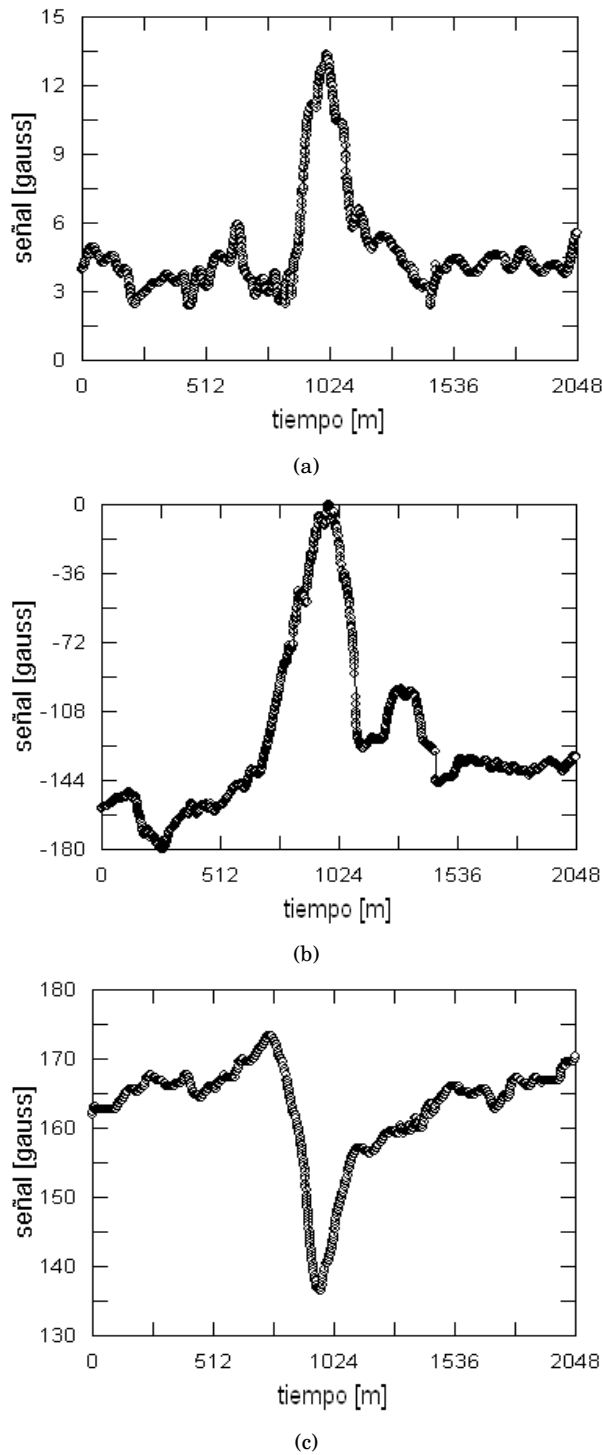


FIG. 1.— (a) Variación de la componente D del campo magnético terrestre. (b) Variación de la componente H del campo magnético terrestre. (c) Variación de la componente Z del campo magnético terrestre. **Datos:** Estación Geomagnética de Patacamaya IIF, CF, FCPN (UMSA). En las tres figuras los datos fueron tomados durante aproximadamente un día y medio de registro continuo.

por datos subrogados (Theiler et al. 1992), es, en este caso, bastante directo: en un ensayo con un nivel de confianza $1 - \alpha = 0.95$ para la hipótesis cero, se requiere $\frac{1}{\alpha} = 20$ muestras (los datos experimentales y 19 registros subrogados).

Para cada una de ellas se calcula un parámetro estadístico probador; si el valor resultante para los datos es muy diferente de los valores obtenidos para las series subrogadas, la hipótesis nula se rechaza al nivel de significación $\alpha = 0.05$.

Las secciones siguientes detallan el procedimiento seguido para los datos de las figs. 1. Se ha calculado también el coeficiente maximal de Lyapunov como criterio comparativo de alinealidad.

La determinación de coeficientes de Liapunov requiere la estimación heurística de zonas planas en las curvas del “factor de estiramiento” la cual -en este caso- afecta seriamente la confiabilidad de los valores obtenidos.

El ensayo posterior por hipótesis nula, usando datos subrogados “tipo Fourier”, permitiría -en principio- dilucidar mejor las particularidades seriales de esos datos. El resultado neto del ensayo, sin embargo, implica la alinealidad de los registros H y D y la alinealidad del registro Z. El valor del coeficiente de predicción alineal para la componente D es lo suficientemente bajo como para dar lugar a una “duda razonable” respecto de la conclusividad del ensayo.

Por ello, en las secciones posteriores, se repite el ensayo de confiabilidad al 0.95 empleando datos subrogados restringidos, que se construyen mediante un proceso de minimización controlada, conocido como Temple Simulado (Metropolis et al. 1953; Kirkpatrick et al. 1983).

2. COEFICIENTES DE LYAPUNOV

Por supuesto, la primera opción en un análisis de dinámica no lineal es el cálculo de los coeficientes de Lyapunov el cual, especialmente si hay comportamiento caótico, será -en principio- el final definitivo del ensayo. Como se sabe, sin embargo, el resultado depende grandemente de la calidad y naturaleza de los datos; por ejemplo, en presencia de ruido aleatorio de amplitud comparable a las perturbaciones caóticas, los algoritmos tienden a perder nitidez.

Para el caso presente, se ha empleado el método de Kantz (1994) que consiste en calcular el factor de estiramiento

$$S(e, m, t) = \left\langle \ln \left(\frac{1}{N_{U_n}} \sum_{x_k \in U_n} |x_{n+t} - x_{k+t}| \right) \right\rangle_n \quad (1)$$

para varias dimensiones de incrustación m y tamaños de vecindario e en el espacio

$$\mathbf{x}_n = (x_{n-(m-1)\tau}, x_{n-(m-2)\tau}, \dots, x_n), \quad (2)$$

con un “retardo” representado por τ .

La rutina de cálculo procede con un algoritmo de búsqueda de vecindades U_n , en ese espacio, variando el tamaño de vecindario $e_{min} \leq e \leq e_{max}$, salvo un factor de reescalamiento, mediante

$$\varepsilon_i = \left(\frac{e_{max}}{e_{min}} \right)^{\frac{1}{(k-1)}}$$

$$e = e_{min} (\varepsilon_i)^i,$$

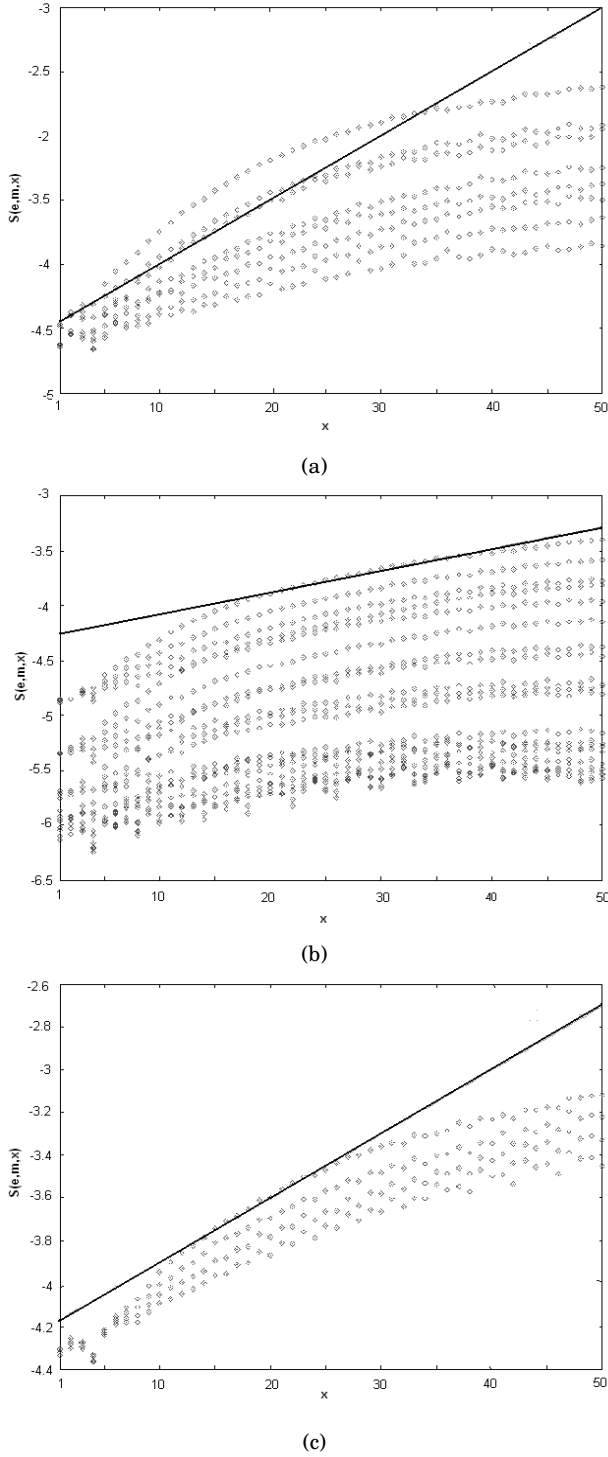


FIG. 2.— Coeficiente (máximo) aproximado de Lyapunov. (a) Registro geomagnético D, la pendiente de la recta es $\lambda = 0.05$. (b) Registro geomagnético H, la pendiente de la recta es $\lambda = 0.02$. (c) Registro geomagnético Z, la pendiente de la recta es $\lambda = 0.03$.

con $0 \leq i < k$.

Los resultados aparecen en las figuras 2(a) 2(b) y 2(c); en ellas, la pendiente de las rectas trazadas en los lugares donde $S(e, m, t)$ muestra tramos -por lo menos aparentemente- lineales.

El rendimiento final del análisis es la presencia de perturbaciones caóticas débiles en las tres com-

ponentes del campo magnético terrestre con coeficientes de Lyapunov maximales $\lambda \approx 0.05$, $\lambda \approx 0.02$ y $\lambda \approx 0.03$ respectivamente.

3. DATOS SUBROGADOS

Sin embargo, para seguir con el esquema planteado en la introducción, sea la hipótesis cero de que los registros geomagnéticos contienen ruido gaussiano filtrado linealmente.

Para el ensayo, a nivel de significación $\alpha = 0.05$, de esa hipótesis cero, se genera $\frac{1}{\alpha} - 1 = 19$ series subrogadas tomando la transformada discreta de Fourier,

$$X_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n \exp\left(\frac{2\pi i k n}{N}\right), \quad (3)$$

multiplicando las amplitudes resultantes por una fase aleatoria y tomando luego la antitransformada:

$$y_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} |X_k| e^{i\beta_k} \exp\left(-\frac{2\pi i k n}{N}\right) \quad 0 \leq \beta_k < 2\pi. \quad (4)$$

Equivalentemente, para fines prácticos, se puede también mezclar aleatoriamente las fases de la transformada y luego tomar la antitransformada.

En las rutinas para lenguaje ensamblado de 32 bites, se ha empleado la transformada rápida de Fourier (TRF) con el algoritmo de Danielson-Lanczos y una adaptación para ese lenguaje del generador de números aleatorios "Mersenne Twister" (Matsumoto & Nishimura 1998).

4. ERROR DE PREDICCIÓN ALINEAL

La propiedad más importante de los datos subrogados, resultantes de implementar las ecuaciones (3) y (4), es la de mantener las propiedades de correlación lineal (pues ésta se calcula mediante la transformada de Fourier) destruyendo, simultáneamente, toda estructura no lineal.

Por tanto, si se usa el error de predicción alineal

$$\gamma(e, m, \tau) = \sqrt{\frac{1}{N} \sum [x_{n+1} - f(x_n)]^2} \quad (5)$$

como parámetro estadístico probador, se espera que el valor correspondiente a los datos experimentales sea menor al de todos los subrogados para rechazar la hipótesis cero; de otro modo, la hipótesis cero es aceptada y se considera que los datos originales sólo contienen ruido gaussiano lineal.

Las funciones $f(x_n)$ en (5) representan, en realidad, un algoritmo de búsqueda de vecindarios a primer orden -descrito en la siguiente sección- del tipo

$$\hat{x}_{n+k} = \frac{1}{N_{U_n}} \sum_{x_j \in U_n} x_{j+k} \quad (6)$$

en el espacio de incrustamiento. Análogamente a (1), en (5) y (6) m es la dimensión de este espacio y e el tamaño de vecindario.

Las figuras 3(a), 3(b) y 3(c) muestran la distribución de valores de γ para el análisis de las tres componentes del campo geomagnético. En ellas, queda claro que la hipótesis cero es rechazable sólo para la componente Z, no obstante que los coeficientes de Lyapunov tienen valores muy comparables para las tres componentes. Obviamente, en este caso, la exigencia en el nivel de significación del ensayo estadístico lo hace más confiable.

5. ALGORITMO DE BÚSQUEDA

Conviene describir brevemente el algoritmo empleado para calcular los valores de (5).

El espacio (2) queda también representado mediante

$$\{\mathbf{x}_i\}_N = \{x_{i,1}, x_{i,2}, \dots, x_{i,m}\}_N.$$

Entonces, para un punto de referencia cualquiera \mathbf{x}_{i_0} , los k vecinos más cercanos son aquellos $\{\mathbf{x}_{i_p}\}_k$ que tienen las distancias más pequeñas respecto de \mathbf{x}_{i_0} (el valor absoluto de la diferencia, aunque pudiera usarse también una distancia euclídea).

Se trata, ahora, de encontrar los índices $\{i_p\}_k$ de los vecinos (evaluando y comparando iterativamente las distancias a todos los puntos del conjunto) y calcular

$$f(x_{i_0}) = \frac{1}{k} \sum_{p=1}^k x_{i_p,1},$$

que es la función que aparece en (5) una vez aplicada la predicción (6).

6. SUBROGADOS RESTRINGIDOS

Para un registro experimental $\{x\}$, es posible generar subrogados $\{y\}$ ejecutando permutaciones que tiendan a minimizar, iterativamente, la función de costo

$$E(\{y_i\}) = \left(\sum_i^K |w_i f(\{y_i\})|^n \right)^{\frac{1}{n}} \quad (7)$$

donde las w_i son factores de peso arbitrarios y $f(\{y_i\})$ es un conjunto de K restricciones

$$f(\{y_j\}) = 0 \quad j = 0, \dots, K. \quad (8)$$

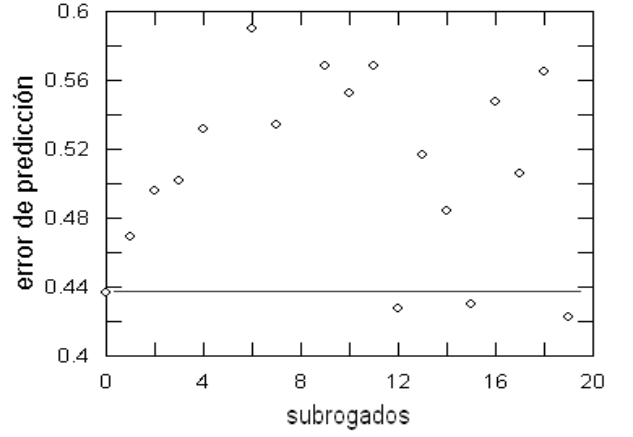
Este proceso de “aleatorización general restringida” rendirá subrogados con las propiedades que se desee, dependiendo de las restricciones elegidas.

Para generar subrogados con las mismas propiedades de linealidad que los datos experimentales, es conveniente introducir, como restricción, la invariancia de la autocorrelación (Schreiber & Schmitz 2000):

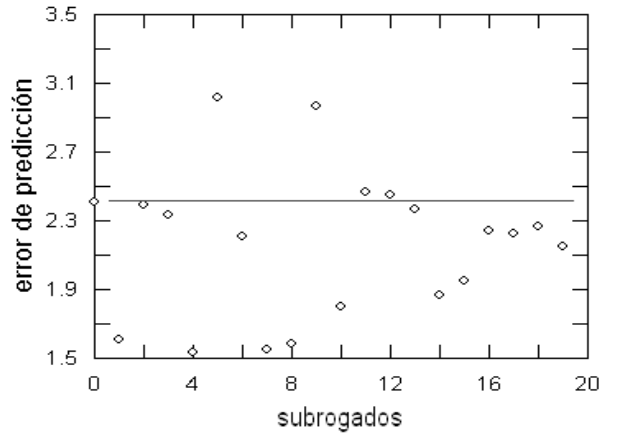
$$A(\{y\}_j) = \sum_{i=0}^{N-1} y_i y_{i+j} = \text{Re}[TRFI(yy^*)] \quad (9)$$

de modo que (8) es ahora

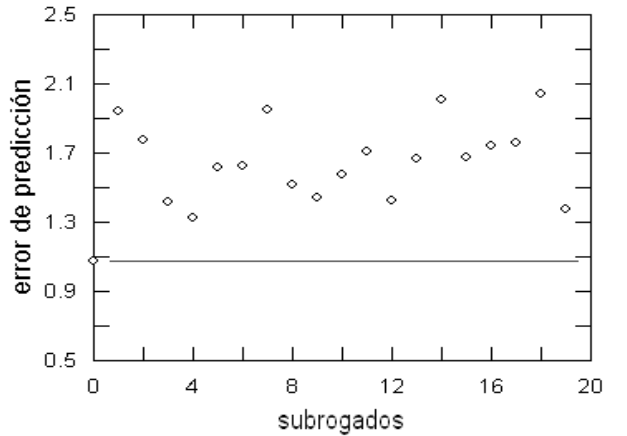
$$A(\{y\}_j) - A(\{x\}_j) = 0 \quad j = 1, \dots, N-1. \quad (10)$$



(a)



(b)



(c)

FIG. 3.— Distribución de valores del error de predicción no lineal para las series subrogadas y para los datos experimentales marcado con una recta horizontal y colocado en el origen de abscisas. (a) Componente D. (b) Componente H. (c) Componente Z.

y, consecuentemente, la función de costo a minimizar es

$$E(\{y\}) = \sum_{j=1}^{N-1} |A(\{y\}_j) - A(\{x\}_j)| \quad (11)$$

donde los $\{y\}$ son permutaciones del original $\{x\}$.

7. TEMPLE SIMULADO

Si se introduce un parámetro T tal que, a un cierto valor T_0 , se asocia un desorden arbitrario en el conjunto de datos $\{x\}$, es posible considerar la función de costo (11) como si fuera la energía de un sistema termodinámico a la temperatura T_0 . A partir de ese punto, se busca la configuración $\{y\}$ para la cual la energía E es mínima a la temperatura final $T_f < T_0$ en un proceso de enfriamiento lento (temple). Siguiendo la distribución de Boltzmann, cada configuración producida por una permutación de dos valores aleatorios debe ocurrir siguiendo el “paso de Metropolis”, esto es, con probabilidad p según:

$$p = 1 \quad \text{si } \Delta E < 0$$

$$p = e^{-\Delta E/T} \quad \text{si } \Delta E \geq 0$$

En la práctica, el paso de Metrópolis se aplica estableciendo, en el curso de cada ciclo, un valor maximal $E_m = E - \Delta E$ con $\Delta E = T \ln(p)$ luego de un sorteo aleatorio de p . Cada ciclo empieza con una permutación aleatoria y concluye con su aceptación (y enfriamiento $T = eT, e < 1$ si $E < E_m$) o rechazo (si $E \geq E_m$), restituyendo, en este caso, los valores permutados.

8. PRUEBA DE ALINEALIDAD

Los datos para las componentes D, H y Z de las figuras (1), contienen tendencias macroscópicas con pequeñas oscilaciones erráticas superpuestas. Como en el caso anterior, la hipótesis cero supone que tales irregularidades consisten de ruido gaussiano filtrado linealmente.

El ensayo, a nivel de significación $\alpha = 0.05$, de esa hipótesis, entonces, requiere también la generación de 19 series subrogadas usando el método descrito en las secciones 6. y 7.

Como los subrogados han sido construidos manteniendo las propiedades de correlación lineal (por la relación entre la transformada de Fourier y la función de autocorrelación), aún es apropiado el uso del error de predicción alineal como parámetro estadístico del ensayo, estimando que el valor correspondiente a los datos experimentales sea menor al de todos los subrogados para rechazar la hipótesis cero.

El error de predicción alineal está dado, entonces, por (5) con las funciones $f(x_n)$ representando también el algoritmo de búsqueda (6) de vecindarios a primer orden en el espacio de incrustamiento (tridimensional, $m = 3$, para todas las pruebas).

9. RESULTADOS

Las figuras 4(a), 4(b) y 4(c) muestran la distribución de valores γ para el análisis de las componentes D, H y Z, respectivamente, del campo geomagnético local. Por ellas, queda claro que la hipótesis cero es rechazable para las tres componentes, esto es, se trata, definitivamente, de tres registros no lineales.

Obviamente, en este caso, la exigencia en el nivel de significación y la notoria separación de los valores

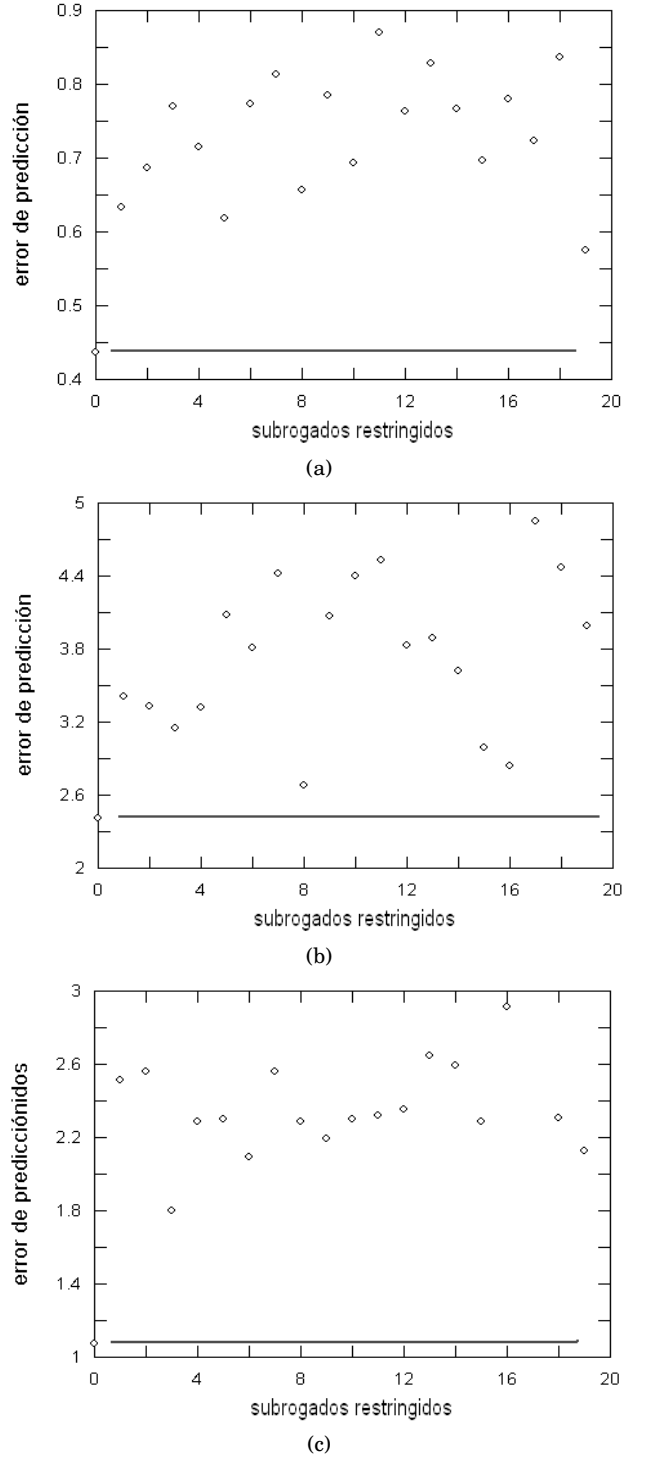


FIG. 4.— Distribución de valores del error de predicción no lineal para las series subrogadas y para los datos experimentales marcado con una recta horizontal y colocado en el origen de abscisas. (a) Componente D. (b) Componente H. (c) Componente Z.

γ entre los datos experimentales y sus subrogados - en los tres casos- dan un ensayo estadístico más confiable y conclusivo.

10. CONCLUSIONES

Siendo los registros D, H y Z simultáneos, es expectable que sus propiedades seriales sean análogas;

por ello, tanto la estimación de los coeficientes máximos de Lyapunov cuanto el ensayo por datos subrogados “tipo Fourier” indican que, excepto para la componente Z, las fluctuaciones estadísticas tienden a ocultar elementos no lineales contenidos de manera menos conspicua en los registros D y H.

El ensayo por subrogados restringidos, por lo visto, es más sofisticado y exigente. Al final, consiste en encontrar las permutaciones que mantengan invariante la autocorrelación, esto es, indirectamente, las propiedades lineales de la serie temporal.

Consecuentemente, los resultados del ensayo, como una inspección de las figuras 4. lo manifiesta, son claramente consistentes y confiables por cuanto:

(i) La dispersión de los valores de γ indica que, ciertamente, el proceso es completamente aleatorio.

(ii) El valor γ es marcadamente inferior al de los subrogados respectivos, especialmente para las componentes D y Z. Y la separación es menor para la componente H, la más dudosa en los otros ensayos.

Por tanto, las tres componentes del campo estudiado resultan ser series temporales no lineales. Obviamente, la naturaleza específica de esa alinealidad debe ser objeto de otros análisis, si resultara ser muy importante.

Por lo demás, al igual que en otras disciplinas (Skinner et al. 1994), las series temporales en Física, cuando el problema de analizar propiedades de linealidad es pertinente, el método por datos subrogados permite: verificar resultados obtenidos por otros métodos (por ejemplo, como se ha visto, el cálculo del coeficiente máximo de Lyapunov) y seleccionar aquellos registros que pueden ser objeto de tratamientos más detallados de alinealidad, comportamiento caótico, etc. y, claro, desechar rápidamente las series experimentales poco interesantes para tales propósitos.

Las rutinas computacionales fueron elaboradas específicamente para los objetivos perseguidos. Aunque, en general, no son de gran sofisticación, el empleo del lenguaje ensamblado, el especial cuidado en la optimización del código y el generador de números aleatorios adecuado (Matsumoto & Nishimura 1998) para el sorteo Monte Carlo, dan excelentes resultados en lo que a velocidad y distribución de valores finales (figs. 3.) se refiere.

Obviamente, aunque no es necesario para este caso específico, tanto el número de iteraciones para la estimación del coeficiente máximo de Lyapunov cuanto el nivel de confianza $1 - \alpha$ del ensayo de hipótesis podrían ser aumentados mejorando, así, la sensibilidad del análisis.

Conviene poner de manifiesto las siguientes tres observaciones:

Primera, la de que los tres métodos fueron ensayados exitosamente con series alineales conocidas (casos especiales de la serie logística y del mapa de Hénon), principalmente para asegurar la efectividad de los algoritmos diseñados para este trabajo.

Segunda, los registros geomagnéticos exhiben cierta periodicidad diaria. El número de datos -en potencias de 2- fue elegido para analizar, aproximadamente, ese lapso. Series más cortas o más largas no son útiles en el empleo práctico de los efectos investigados.

Finalmente, en referencia a los problemas que plantea la presencia de no estacionariedad en los datos (mencionados, por ejemplo, en Theiler et al. (1992)), se ha puesto especial cuidado en que los subrogados no alteren significativamente el comportamiento temporal de los datos, de manera que la hipótesis nula no sea invalidada; el tamaño de los registros es también circunstancialmente útil a este respecto.

REFERENCIAS

- Kantz H. (1994), *Phys. Lett. A* **185**, 77
 Kirkpatrick S., Gelatt Jr. C. D. & Vecchi M. P. (1983), *Science* **220**, 671
 Matsumoto M. & Nishimura T. (1998), *ACM Transactions on Modeling and Computer Simulation* **8**, 3
 Metropolis N., Rosenbluth A., Rosenbluth M., Teller A. & Teller E. (1953), *J. Chem. Phys.* **21**, 1097
 Ricaldi E. (2012), *Datos del OBSERVATORIO GEOMAGNÉTICO PATACAMAYA, IIF-UMSA, La Paz, Bolivia* (Comunicación Privada)
 Schreiber T. & Schmitz A. (2000), *Physica D* **142**, 346
 Skinner J. E., Molnar M. & Tomberg C. (1994), *Integ. Physiol. Behavior. Sci.* **29**, 217
 Theiler J., Eubank S., Longtin A., Galdrikian B. & Farmer J. D. (1992), *Physica D* **58**, 77