

Detección de Estacionalidad en Series de Tiempo

E. A. Santos-Camacho, J. G. Figueroa-Nazuno

Centro de Investigación en Computación (CIC)
araceli.libelula@gmail.com, jfn@cic.ipn.mx

Resumen. El análisis de series de tiempo tiene como objetivo investigar el comportamiento de una o más variables $t(x)$. Los ejemplos de su aplicación son muchos y muy variados como tasa de inflación, índice de precios, meteorología, contaminación, tasa de natalidad y el efecto de El Niño, por mencionar algunos. Usando técnicas modernas en el análisis de series de tiempo se pueden detectar patrones en su comportamiento, pero en muchos casos una de las técnicas más usadas es la Transformada Rápida de Fourier (FFT por sus siglas en inglés); esta técnica es muy rígida y hace inadecuado su uso en series de tiempo naturales. En este trabajo se presenta una técnica que nos permite aportar un mayor conocimiento sobre la estacionalidad en los fenómenos naturales. Los resultados experimentales indican que la técnica de Lomb-Scargle es una alternativa para detectar estacionalidad empleando criterios matemáticos y estadísticos.

Palabras clave: Estacionalidad, serie de tiempo, periodograma Lomb-Scargle, Transformada Rápida de Fourier.

1 Introducción

Una forma de representar fenómenos naturales es mediante una serie numérica de tiempo, la cual puede verse como un conjunto de valores en forma consecutiva en el tiempo. Empleando diferentes técnicas de análisis de señales podemos describir el comportamiento del fenómeno a corto y mediano plazo. Sin embargo, es muy difícil determinar las diferentes características que pueden presentar [1] debido a que comúnmente las series de tiempo no presentan un muestreo regular y/o no son uniformes, lo cual genera problemas en la búsqueda de características, en especial la búsqueda de estacionalidad, limitando el uso de técnicas como FFT, ya que el uso de este tipo de técnicas son muy estrictas.

La estacionalidad es una característica o patrón repetitivo en los fenómenos y puede presentarse en períodos anuales, mensuales, semanales y otros diarios; pero en muchas ocasiones estos patrones se presentan en forma irregular para el mismo fenómeno.

no por lo que esto complica su estudio. Como una alternativa se ha propuesto la técnica de Lomb-Scargle (1976 y 1982), el cual es un método de estimación de espectro de frecuencias, sobre la base de mínimos cuadrados ajustando ondas sinusoides a muestras de datos, en donde se toma primero un retardo de tiempo τ , logrando obtener senosoides mutuamente ortogonales en tiempos de muestreo t_i , para obtener la mejor estimación en la frecuencia de poder [2,3], y posteriormente usar técnicas estadísticas, para la obtención de estacionalidad en series de tiempo.

2 Serie de tiempo

Una serie de tiempo, es un conjunto de datos numéricos obtenidos a partir de una observación experimental de algún sistema o mediante el cálculo numérico de ecuaciones, es decir, una serie de tiempo es un conjunto de tipo:

$$ST = \{x_1, x_2, \dots, x_t, \dots, x_N\}$$

Una serie de tiempo por lo general consta de diferentes oscilaciones de valores, sin embargo, las series de tiempo naturales raramente tienen oscilaciones tan simples como las que se muestran en la figura 1, por lo que tienden a ser más complejas, como se muestra en la figura 2 [4].

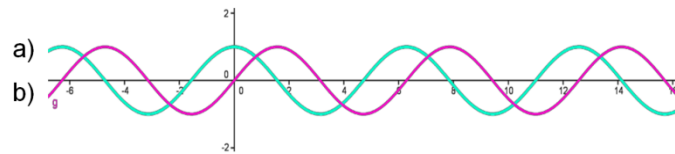


Fig. 1. Ejemplo de a) serie de tiempo seno y b) serie de tiempo coseno

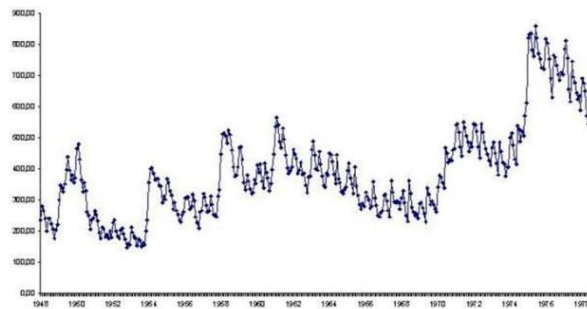


Fig. 2. Ejemplo de serie de tiempo natural.

3 Técnica lomb-scargle

Un método para extraer y evaluar una serie de tiempo periódica y obtener de forma significativa estacionalidad en el sentido estadístico con datos irregularmente espa-

ciados, es la técnica Lomb-Scargle [5], la cual evalúa los datos como senos y cosenos solo en el tiempo t_i que se está midiendo en ese momento, como se describe a continuación:

Dados N mediciones $h_i \equiv h(t_i)$ para $i = 0$ hasta $N - 1$, se puede calcular el periodograma mediante la ecuación (1) [6].

$$P_{\omega}(f) = \frac{1}{2\sigma^2} \left\{ \frac{[\sum (h_i - \bar{h}) \cos \omega(t_i - \tau)]^2}{\sum \cos^2 \omega(t_i - \tau)} + \frac{[\sum (h_i - \bar{h}) \sin \omega(t_i - \tau)]^2}{\sum \sin^2 \omega(t_i - \tau)} \right\} \quad (1)$$

donde $\bar{h} = \frac{1}{N} \sum_{i=0}^{N-1} h_i$, $\sigma^2 \equiv \frac{1}{N-1} \sum_{i=0}^{N-1} (h_i - \bar{h})^2$, $\omega \equiv 2\pi f > 0$ y τ es definida por la

$$\text{relación } \tan(2\omega\tau) = \frac{\sum \sin 2\omega t_j}{\sum \cos 2\omega t_j}.$$

La distribución de probabilidad para datos aleatorios (no periódicos) es $\exp(-P_{\omega})$. Si $P \cong 3$, existe $\sim 5\%$ de probabilidad de que el componente sea aleatorio. Por otro lado si la prueba se realiza en muchas frecuencias, la probabilidad “máxima” aleatoria se expresa en la ecuación (2):

$$P_{\omega} \equiv 1 - (1 - e^{-P_{\omega}})^M \quad (2)$$

Donde M es el número de frecuencias, $\sim 1.2x$ el número de observaciones.

3.1 Pseudocódigo

A continuación se presenta el pseudocódigo, para calcular estacionalidad estadística con la técnica de Lomb- Scargle para datos irregularmente espaciados:

Entradas: N (serie de tiempo), T (intervalos de tiempo), $hifac \geq 1$ y $ofac \geq 4$.

Inicio

```

    inicializar N  $\leftarrow$  longitud de h;
    inicializar m1  $\leftarrow$  el máximo valor de t;
    inicializar m2  $\leftarrow$  el mínimo valor de t;
    inicializar Suma  $\leftarrow$  0;
    asignar T  $\leftarrow$  m1 - m2;
    para i=0 hasta N-1
        asignar suma  $\leftarrow$  suma + h(i);
    fin
    asignar media  $\leftarrow$  suma / N;
    asignar suma  $\leftarrow$  0;
    para i=0 hasta N-1
        asignar suma  $\leftarrow$  suma + (h(i) - media) ^ 2;
    fin

```

```

asignar varianza  $\leftarrow$  suma / (N - 1);
asignar j  $\leftarrow$  0;
para i  $\leftarrow$  1/(T*ofac) hasta hifac*N/(2*T)), en incrementos de 1/(T*ofac)
    asignar f[j]  $\leftarrow$  i;
    asignar j  $\leftarrow$  j+1;
fin
para i  $\leftarrow$  0 hasta longitud de f-1
    asignar w[i]  $\leftarrow$  f[i] * 2 *  $\pi$ ;
fin
para i=0 hasta longitud de w-1
    asignar tau[i]  $\leftarrow$   $\arccotang\left(\frac{\text{seno}(2*w[1]*t[1])}{2*w[i]}\right), \left(\frac{\text{coseno}(2*w[1]*t[1])}{2*w[i]}\right)$ 
fin
para i  $\leftarrow$  0 hasta longitud de w-1
    asignar Asin[i]  $\leftarrow$   $\text{coseno}(w[i] * t[i] - \text{repmat}(w[1] * \text{tau}[1], \text{longitud de } t))$ ;
    asignar Bcos[i]  $\leftarrow$   $\text{seno}(w[i] * t[i] - \text{repmat}(w[1] * \text{tau}[1], \text{longitud de } t))$ ;
fin
para
    asignar P  $\leftarrow$   $\text{sumar}(Bcos * \text{diagonal}(h - \mu))$ 
fin
asignar M  $\leftarrow$  (2 * longitud de f)/ofac
asignar prob  $\leftarrow$  M +  $\exp(-P)$ ;
asignar inds  $\leftarrow$  prob > 0.01;
asignar prob(inds)  $\leftarrow$   $1 - \left(1 - \exp(-P(\text{inds}))\right)^M$ ;
Fin

```

4 Metodología

Dado las bases de datos con series de tiempo naturales y artificiales se les aplica la técnica de Lomb-Scargle empleando la ecuación 1, para calcular el periodograma y posteriormente se hace el análisis estadístico, obteniendo como resultado dos graficas; en la figura 3, se presenta esta descripción de forma visual.

En la figura 4, se presenta una grafica en la cual el eje de las “x” representa la frecuencia. Para la figura izquierda se presenta la frecuencia de la serie de tiempo obtenida a partir de la técnica Lomb-Scargle. En la figura derecha el eje de las “y” representa el grado de significancia. En este ejemplo todas las frecuencias que sobrepasan el umbral de significancia marcado con el numero 1 en la grafica izquierda, se filtran para obtener la grafica derecha como se muestra en el numero 2.

1. Se obtiene el conjunto de datos.
 2. Se aplica la ecuación (1) perteneciente a la técnica de Lomb-Scargle.
 3. Se obtiene la estacionalidad estadística.
- Nota. La significancia es un parámetro.

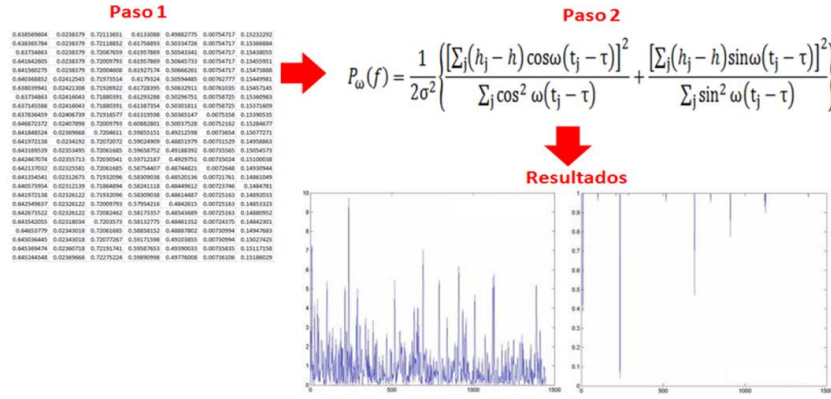


Fig. 3. Metodología para la extracción de estacionalidad empleando la técnica Lomb-Scargle.

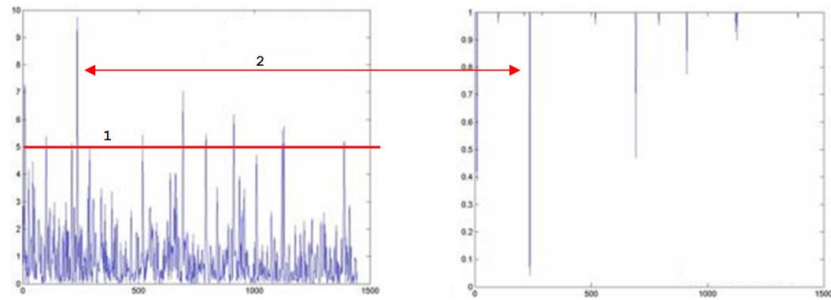


Fig. 4. En la parte izquierda se presenta la frecuencia obtenida a partir de una serie de tiempo, y en la parte derecha se presenta la frecuencia significativa, la cual sobrepasa el umbral marcado con la línea numero 1.

5 Conjuntos de datos

Para el análisis experimental se tomaron dos conjuntos de datos, el conjunto A contiene series de tiempo naturales, las cuales han sido muy estudiadas y se sabe que presentan estacionalidad [7]. El segundo conjunto B, contiene series artificiales en los cuales no se sabe si presentan o no estacionalidad.

5.1 Conjunto de datos A (datos naturales)

Para el análisis experimental con series naturales, se tomaron los datos de 8 índices de telefonía [8, 9], cuyas mediciones fueron tomadas desde 1950 hasta 2010, para cada uno. A estos datos se les realizó una transformación lineal [10]. En la tabla 1, se muestran las siglas utilizadas para identificar los en la descripción de resultados.

Table 1. Índices de teleconexión.

Siglas	Índices de teleconexión	Siglas	Índices de teleconexión
NOA	North Atlantic Pattern	PNA	Pacific North American Pattern
EA	East Pacific Pattern	EA/WR	East Atlantic/West Russia Pattern
WP	West Pasific Pattern	SCA	Scandinavia Pattern
EP/NP	East Pasific/North Pacifica Pattern	POL	Polar Eurasia Pattern

5.2 Conjunto de datos B (datos sintéticos)

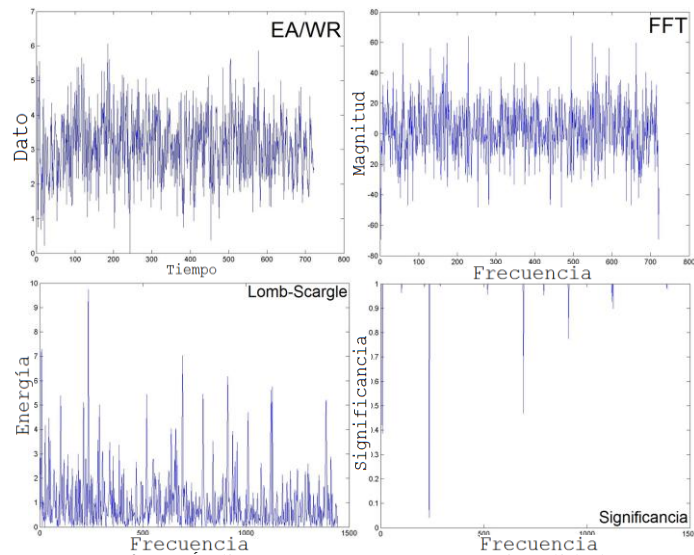
Se tomaron 30 series de tiempo obtenidas a partir de ecuaciones y análisis experimental, en la tabla 2 se enlista el nombre de cada una de ellas.

Table 2. Series sintéticas.

Santa Fe	DownJones	HIVDNA	Logistic	Primos	Sine
SCII	ECG	HumanDNA	Lorenz	QP2	Star
BrownMotion	EEG	IKEDA	Lovaina	QP3	Tent
Cantor	El Niño	Kobe	Mackey-Glass	Rossier	Vaderpol
D1	Hennon	Laser	Plasma	S&P500	White Noise

6 Análisis experimental

A cada serie de tiempo de ambos conjuntos de datos se les aplicó la técnica de Lomb-Scargle y la FFT, para obtener el periodograma de cada una de ella, algunos de los resultados muestran a continuación.

**Fig. 5.** Índice EA/WR presenta estacionalidad.

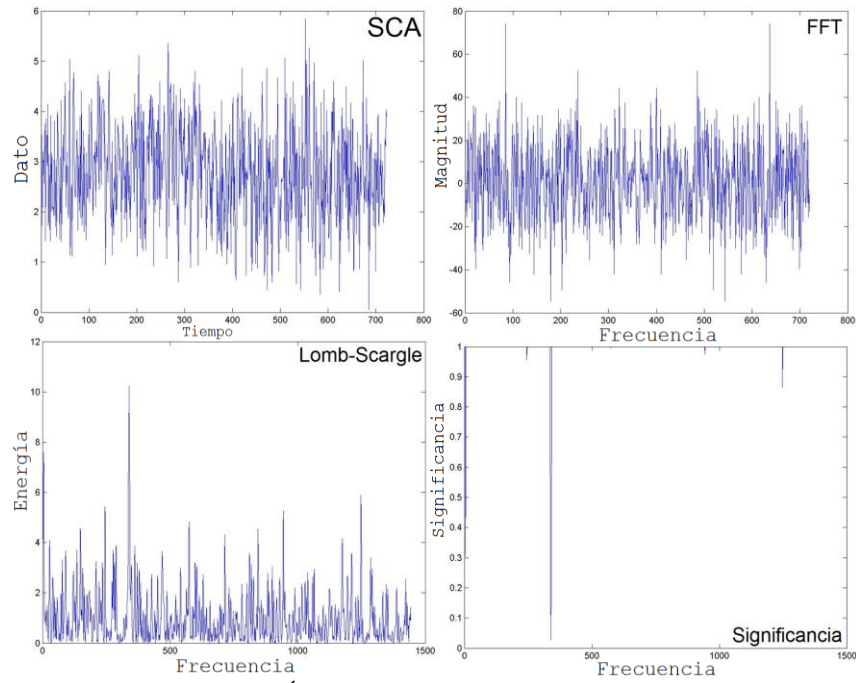


Fig. 6. Índice SCA, presenta estacionalidad.

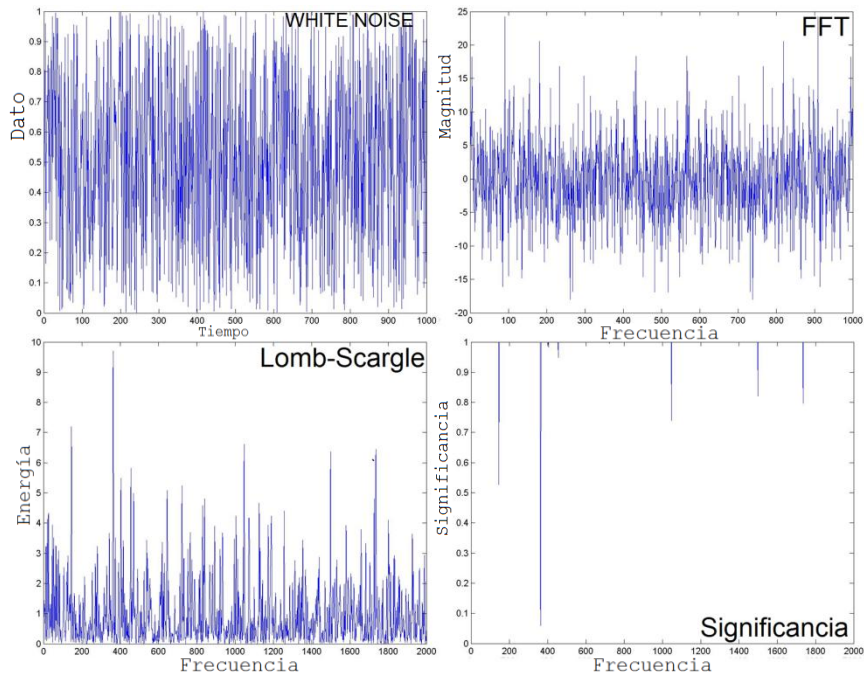


Fig. 7. Serie White Noise, presenta estacionalidad.

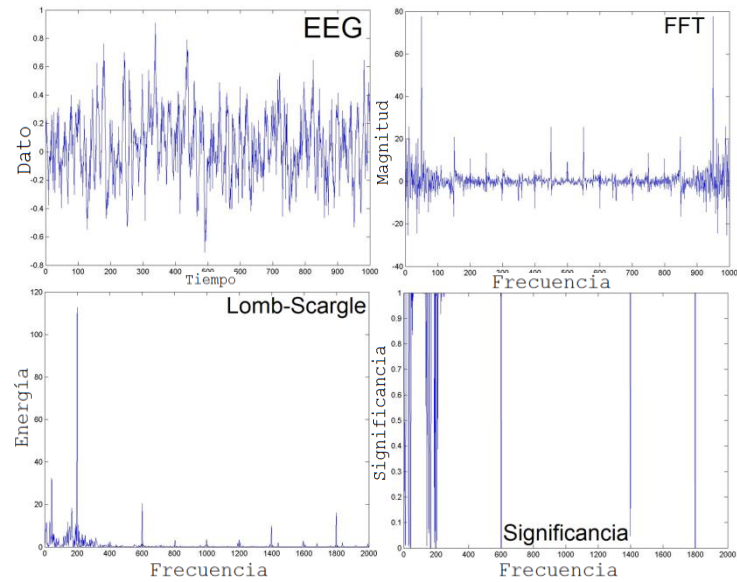


Fig. 8. Serie EEG, presenta estacionalidad.

Como se puede observar en los diferentes ejemplos mostrados en las figuras 4 - 8, las series presentan estacionalidad observada a través de la técnica de Lomb-Scargle, además de obtener la estacionalidad significativa. Por otra parte, las series de la figuras 9 y 10 para el conjunto de índices y el conjunto B respectivamente no presentan estacionalidad, por lo que el comportamiento que presentan las series de tiempo es complejo.

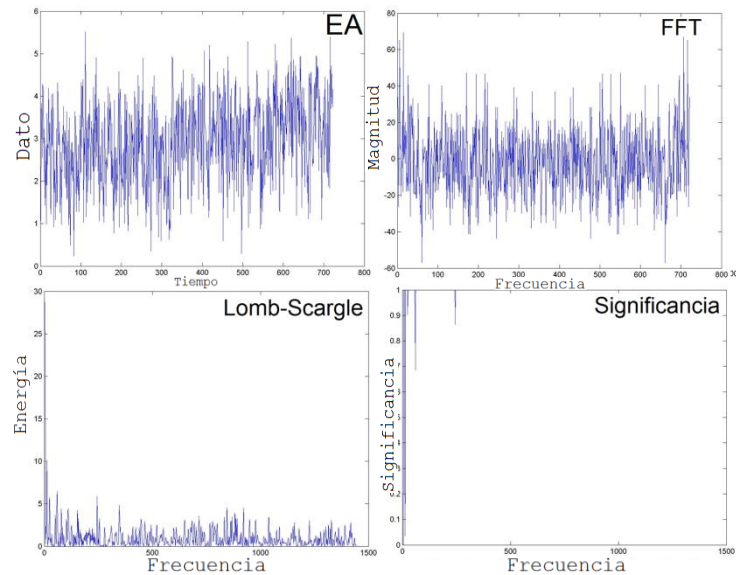


Fig. 9. Índice EA, no presenta estacionalidad.

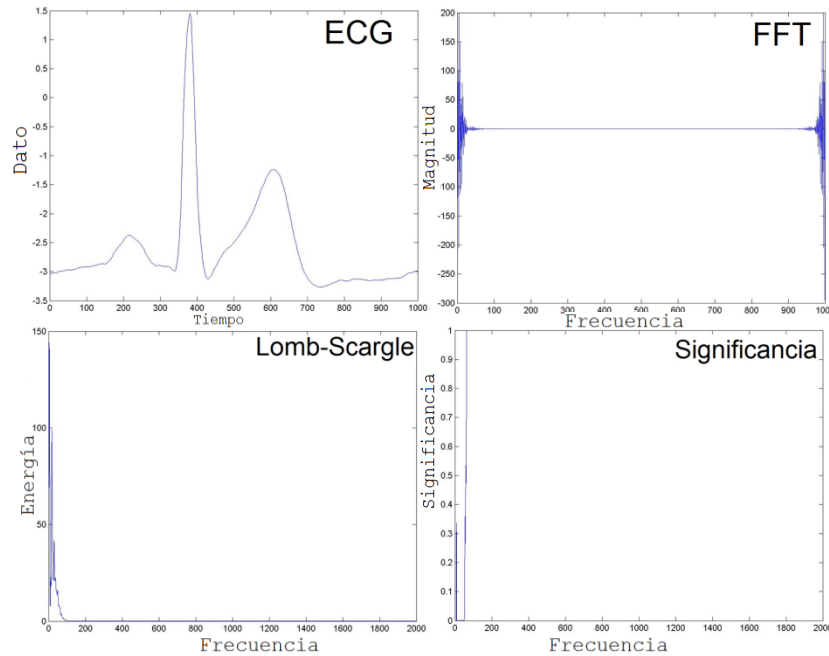


Fig. 10. Serie ECG, no presenta estacionalidad.

7 Conclusiones

Uno de los paradigmas que está latente y sigue siendo un reto en diferentes áreas es predecir fenómenos naturales, por lo que una forma de atacarlo es conociendo el comportamiento de las series de tiempo, actualmente una de las características más significativas es la estacionalidad, pero la mayor parte de las series presentan un grado de dispersión variable, es decir, se presenta el mismo fenómeno pero no con el mismo rango de tiempo por lo que al manifestarse de esta forma no es claro conocer su dinámica, sin embargo, existen técnicas que pueden mitigar esta condición.

En este trabajo se presenta el análisis de series naturales y artificiales, en los que se presenta estacionalidad estadísticamente significativa, sin embargo, empleando la técnica FFT, no es claro detectarla como se puede observar en los diferentes ejemplos debido al funcionamiento de técnica de FFT, ya que para su análisis requiere que sean acotadas, infinitas, por mencionar algunas características y esto no es posible encontrarlo en series naturales, por lo que al aplicar FFT se puede llegar a tener resultados erróneos. Los resultados obtenidos muestran el potencial de la técnica Lomb-Scargle, para la obtención de estacionalidad en las series de tiempo natural y artificial, empleando métodos estadísticos que con otras técnicas no se puede hacer.

Referencias

1. E. F. Bautista Thompson, "Medición de la predictibilidad de series de tiempo: un estudio experimental, "Tesis de Doctorado en Ciencias de la Computación del Centro de Investigación en Computación del Instituto Politécnico Nacional, México, 2005.
2. Press, "Numerical Recipes", Cambridge University Press, 3rd ed., 2007.
3. J. D. Scargle, "Studies in astronomical time series analysis. II – Statistical aspects of spectral analysis of unevenly spaced data", *Astrophysical Journal* 263: 835, 1982.
4. G. P. Weedon, "Time-Series Analysis and Cyclostratigraphy", Cambridge University Press, 2003.
5. R. H. D. Townsend, "Fast Calculation of the Lomb–Scargle Periodogram using Graphics Processing Units", *The Astrophysical Journal Supplement Series*, 191:247– 253, The American Astronomical Society, 2010.
6. N. Lomb, "Least-Squares Frequency Analysis of Unqually Spaced Data", *Astrophysics and Space Science*, 39:447-462, 1975.
7. E. Bautista-Thompson, E. Guzmán-Ramírez y J. Figueroa-Nazuno, "Predicción de Múltiples Puntos de Series de Tiempo Utilizando Support Vector Machines", *Computación y Sistemas*, vol. 7 No. 3 pp. 148-155, México 2004.
8. J. W. Hurrell, Y. Kushnir, G. Ottersen, M. Visbeck, "The North Atlantic Oscillation: climatic significance and environmental impact.", *Geophysical Monograph Series*. Washington, DC: American Geophysical Union, 2003.
9. J. W. Hurrell, "Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation", *Science*, vol. 169, pp. 676-679, 1995.
10. E. A. Santos-Camacho & J. Figueroa-Nazuno, "Clasificación de Series de Tiempo Mediante una Comparación Elástica", Tesis que para obtener el grado de Licenciatura, Universidad Autónoma del Estado de México, 2012.