

COMPLEJIDAD ESTADÍSTICA EN SERIES TEMPORALES: APLICACIÓN A SEÑALES EEG

(Statistical complexity in time series: Application to EEG signals)

¹Escalona-Morán, M.; ²Molina, L. A.; ²Cosenza, M. G.

¹Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB),
Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, España,

²Centro de Física Fundamental, Universidad de los Andes, Mérida, Venezuela,
Correo Electrónico: miguelangel@ifisc.uib-csic.es

RESUMEN

Se presenta un método para calcular la complejidad estadística en series de tiempo haciendo uso de la noción de complejidad introducida por López-Ruiz *et al.* en 1995. Se aplica esta medida de complejidad a series de tiempo de sistemas dinámicos no lineales. La utilidad de este método se muestra en el análisis de señales electroencefalográficas. Se obtiene la evolución temporal de la complejidad de registros electroencefalográficos de un paciente epiléptico y de un sujeto sano y se muestra que la patología epiléptica está asociada con un nivel de complejidad menor en comparación con la que corresponde a un sujeto sano.

Palabras Clave: Complejidad, análisis de series de tiempo, señales EEG

ABSTRACT

We present a method to compute the statistical complexity of time series by using the notion of complexity introduced by Lopez-Ruiz *et al.* in 1995. This measure of complexity is applied to time series from nonlinear dynamical systems. The usefulness of this method is shown in the analysis of electroencephalographic records from an epileptic patient and from a healthy subject. By the use of this method, it is constructed the time evolution of the complexity of electroencephalographic records showing that the epileptical pathology is associated to a lower level of complexity in comparison to that corresponding to a healthy subject.

Key Words: Complexity, time series analysis, EEG signals

INTRODUCCIÓN

El estudio de las propiedades macroscópicas emergentes, tanto espaciales como temporales, en sistemas de elementos dinámicos no lineales que interactúan entre sí, es un foco de mucha interés en la literatura científica reciente. La observación de comportamientos colectivos ordenados que surgen en una gran variedad de estos sistemas indican que sus comportamientos son, en general, no triviales y que además presentan características universales, independientemente del contexto, las cuales no pueden derivarse a partir del conocimiento de las propiedades de los elementos constituyentes. Fenómenos como la formación espontánea de estructuras, patrones espaciales, sincronización y oscilaciones colectivas, ondas espirales, segregación y diferenciación, y crecimiento de dominios y de interfaces, son ejemplos de procesos de auto-organización que ocurren en diversos sistemas, tales como sistemas físicos, químicos, biológicos, sociales, económicos, etc. Estos efectos colectivos son el resultado de la interacción y cooperación entre los elementos constituyentes de un sistema. El concepto general de sistemas complejos se ha aplicado a estos conjuntos de elementos capaces de generar estructuras o funciones globales ausentes a nivel local.

Los fenómenos emergentes en sistemas complejos pueden exhibir propiedades universales, independientemente de los detalles del sustrato en el cual éstos se expresan. Consideremos, por ejemplo, las reglas que gobiernan las operaciones aritméticas elementales. Estas reglas no varían dependiendo del sistema físico (un computador) o biológico (un cerebro) en el cual se manifiestan, y por lo tanto son lógicamente independientes de las leyes físicas que los rigen. Igualmente, un fluido turbulento y un cerebro son claramente muy diferentes a nivel microscópico, pero ambos comparten un rasgo importante: presentan la imposibilidad de predecir la riqueza de sus dinámicas por mera extrapolación del comportamiento de sus unidades básicas.

En este estudio se hace uso de herramientas de la dinámica no lineal para investigar la evolución de los niveles de complejidad en series temporales. Diversas definiciones operativas de complejidad han sido propuestas para cuantificar el grado de complejidad de un sistema (Kolmogorov, 1965; Chaitin, 1990; Lempel, y Ziv, 1976; Bennett, 1988; Grassberger, 1986), dependiendo de los aspectos colectivos o funcionales que se intentan caracterizar. En los últimos años, (López-Ruiz *et al.* 1995) propusieron una medida de

complejidad estadística que refleja la complejidad estructural de un sistema y que resulta sencilla de calcular en la mayoría de los casos. La noción de complejidad LMC ofrece un novedoso punto de vista basado en la descripción estadística de los sistemas a una escala dada. En este esquema, el conocimiento de las leyes físicas que gobiernan la evolución dinámica en la escala específica debe ser usado para encontrar los estados accesibles de un sistema y su distribución de probabilidad. Este proceso indicaría el valor de la complejidad. En esencia, la complejidad LMC es una interpolación entre la información almacenada por el sistema y la distancia a la equipartición (medida de una jerarquía probabilística entre las partes observadas) de la distribución de probabilidad de sus estados accesibles. Su principal cualidad consiste en proveer un concepto intuitivo de complejidad, el cual ha demostrado ser satisfactorio para discernir situaciones reconocidas como complejas.

MÉTODO

Medida de complejidad LMC

Asumamos que a una cierta escala de observación, o nivel de descripción, un sistema posee N estados accesibles $\{x_1, x_2, \dots, x_N\}$ y una distribución de probabilidad correspondiente $\{p_1, p_2, \dots, p_N\}$. (Shannon y Weaver, 1949) demostraron, bajo las más elementales normas de consistencia, que la única función que da cuenta de la información almacenada en tal sistema es

$$H = -k \sum_{i=1}^N p_i \log p_i \quad (1)$$

donde k es una constante. Es sencillo determinar que la información H contenida en un cristal ordenado es $H_{cristal} \approx 0$ mientras que un gas aislado térmicamente tiene una distribución equiprobable, $p_i = 1/N$, y entonces $H_{gas} \approx -k \log N$; lo cual representa la máxima información posible para un sistema de N estados. Cualquier otro sistema tendrá una cantidad de información intermedia entre estos extremos de orden y desorden.

López-Ruiz, *et al.* 1995 introdujeron el concepto de desequilibrio de un sistema como una medida de la distancia de su distribución de probabilidad a la distribución equiprobable, definida como

$$D = \sum_{i=1}^N \left(p_i - \frac{1}{N} \right)^2 \quad (2)$$

Esta función de desequilibrio D es máxima para un cristal perfecto y se anula para un gas ideal. Se ha demostrado que en sistemas aislados fuera de equilibrio, la función D decrece con el tiempo (Perakh 2004).

La complejidad de un sistema resulta de la competencia entre estas dos cantidades. López-Ruiz *et al.* 1995, definen la complejidad estadística de un sistema como el producto de la cantidad de información almacenada en el sistema y su desequilibrio,

$$C = H \cdot D \quad (3)$$

$$= -k \sum_{i=1}^N p_i \log p_i \times \sum_{i=1}^N \left(p_i - \frac{1}{N}\right)^2$$

Esta definición concuerda con las propiedades que, de manera intuitiva, debería poseer la caracterización de la complejidad de un sistema; ya que $C \rightarrow 0$ para sistemas ubicados en los extremos de orden o de desorden (cristal perfecto o gas ideal), mientras que cualquier otro sistema con comportamiento intermedio tendrá $C > 0$. En esencia, la complejidad LMC es una interpolación entre la información almacenada en el sistema y su distancia a la equipartición de la distribución de probabilidad de sus estados accesibles. Esta medida de complejidad se ha calculado analíticamente para varios modelos de sistemas físicos, tanto con distribuciones de probabilidad continuas como discretas (Calbet *et al.* 2001).

Nótese que el valor de la complejidad LMC de un sistema depende de la escala de observación, la cual determina el número de estados observables del sistema. Por ejemplo, una célula observada desde una distancia muy grande puede describirse como un punto, el cual posee complejidad cero; sin embargo, a una distancia adecuada se pueden apreciar estructuras internas que producen comportamientos colectivos emergentes, por lo tanto, la complejidad asociada a este sistema tendrá un valor mayor que cero. Adicionalmente, debe tenerse en cuenta que los valores de la complejidad LMC están definidos con un factor de escala dado por la constante k . Por lo tanto, lo que tiene relevancia para caracterizar un sistema son los valores relativos de la complejidad LMC.

Procesamiento de series temporales

Una serie de tiempo está dada por un conjunto de observaciones ordenadas cronológicamente, que representan los cambios de una variable. Dicha variable

puede ser el resultado de un gran número de procesos que se llevan a cabo dentro de un sistema. Ejemplos de este tipo de variables los encontramos en economía, física, química, biología, medicina, etc.

Podemos clasificar las variables de dos maneras, activas y pasivas. Las variables activas son aquellas que forzan al sistema a cambiar su estado y por tanto juegan un rol importante en el desempeño del sistema bajo estudio, este tipo de variable se conoce también como parámetro. Las variables pasivas son aquellas que dependen de otras variables o parámetros, por lo que su cambio es el resultado de algún otro proceso dentro del sistema. Los mapas iterativos son un ejemplo de ambas variables. Una variable pasiva se hace iterar bajo el dominio de una variable activa o parámetro; un cambio en el valor del parámetro puede derivar en un cambio en el comportamiento de la variable pasiva.

A continuación se presenta un método para calcular los niveles de complejidad en un sistema con ambos tipos de variables. Para poder realizar la medida de complejidad LMC en una serie de tiempo, es necesario:

1. Definir el número de estados observables N del sistema, el cual determina la escala de observación.
2. Conocer la distribución de probabilidades del sistema para el número de estados o escala dada.
3. Calcular la complejidad por medio de la Ec. 3. Este esquema dará como resultado el valor de la complejidad para toda la serie de tiempo, lo que es interesante en casos donde se tiene una gran cantidad de series de tiempo con diferentes valores de parámetros. Tal es el caso de los diagramas de bifurcación (Figura 1), en el que existe una serie de tiempo para cada valor de parámetro.

Si se desea conocer la evolución de los niveles de complejidad a lo largo del registro, podemos realizar el mismo proceso en fracciones sucesivas del registro, lo que añade un nuevo parámetro: el tamaño de la ventana de tiempo tv donde se realizará el cálculo.

El conocimiento de las leyes físicas que rigen el sistema bajo estudio nos ayuda a estimar los parámetros necesarios para realizar el cálculo de la complejidad. Sin embargo, en algunos sistemas, es posible que tengamos que realizar estudios adicionales para determinar estos parámetros. Existe una forma sencilla de obtener un valor aproximado de estos parámetros.

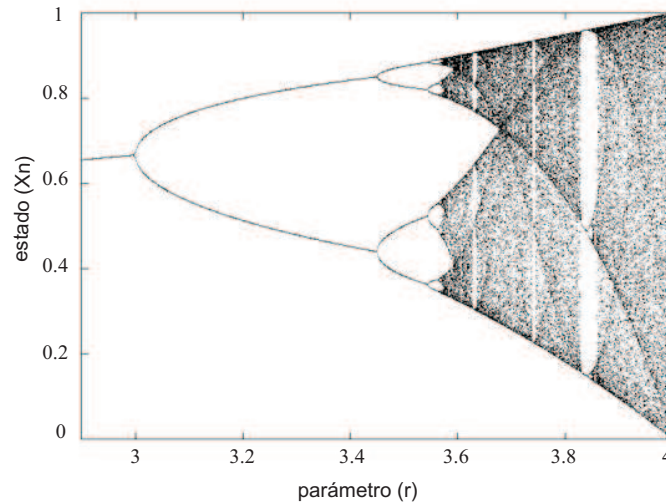


Figura 1. Diagrama de bifurcación del mapa logístico.

1. Número de estados N . El número de estados se refiere a la cantidad de estados que pueden ser visitados por algún valor de la serie de tiempo en su totalidad. Este número depende de la escala de observación, por ejemplo, si $N = 1$, el sistema se está observando a una distancia muy grande, en la que no se nota detalle alguno, por el contrario si $N \rightarrow \infty$ no se logrará observar la dinámica del sistema y la susceptibilidad al ruido será muy alta. En nuestro caso, un gráfico de la complejidad en función de N bastará para definir un valor.
2. Tamaño de la ventana tv . El tamaño de la ventana corresponde al intervalo de tiempo utilizado para realizar una medida de complejidad. Éste intervalo debe ser lo suficientemente grande para incluir la dinámica del sistema.

Aplicación a señales EEG

El cerebro humano está compuesto por más de 1010 neuronas, cada una recibe impulsos eléctricos (conocidos como potenciales de acción) de las otras neuronas por medio de una sinapsis y envía señales de salida a un número similar de neuronas (Shatz, 1981). Cuando una red de neuronas se encuentra activa, ésta produce una diferencia de potencial que puede ser capturada por un electroencefalograma (EEG). Las señales EEG representan el registro de la actividad neurológica como una función del tiempo. La estructura del EEG representa la actividad entre grandes grupos de neuronas, lo que convierte al EEG en una variable pasiva que contiene información de una gran cantidad de

elementos básicos (neuronas). Muchos investigadores han utilizado métodos de la dinámica no lineal en las aplicaciones médicas. Esto se debe a la habilidad de explicar algunos mecanismos no evidentes de la función cerebral (Babloyantz y Destexhe 1986). La epilepsia es un desorden cerebral caracterizado clínicamente por perturbaciones de la función cerebral que a pesar de ser de corto tiempo, son recurrentes. Estos episodios de perturbación son conocidos como crisis epilépticas y pueden estar asociados con pérdida de la conciencia o un comportamiento anormal. Algunos estudios han mostrado la presencia de no linealidad en señales EEG adquiridas en pacientes epilépticos. Casdagli *et al.* 1996 reportó la presencia de no linealidad en registros EEG invasivos en regiones epilépticas de pacientes patológicos. (Palus, 1996), (Thelier y Rapp, 1996) mostraron la presencia de dinámicas no lineales en los EEG de sujetos normales.

Base de datos

El registro de las señales EEG se realizó en 19 canales a partir de electrodos adheridos a la superficie de la cabeza del paciente con pasta conductora, en los sitios establecidos por el sistema internacional 10-20, midiéndose la diferencia de potencial respecto a una referencia constituida por ambas orejas cortocircuitadas. Se digitalizó la señal a una frecuencia de muestreo de 256 Hz y conversión A/D de 12 bits, filtrada digitalmente entre 0,5 y 30 Hz. Se seleccionaron para el análisis, registros de tiempo variable de actividad en reposo de un sujeto sano y un paciente epiléptico

cuyo registro incluye algunos instantes antes y durante una crisis convulsiva espontánea. En la figura 6a se muestra el canal T4-Ref del EEG de un sujeto sano y en la figura 6c se muestra el canal C4-Ref del EEG de un paciente epiléptico antes y durante una crisis convulsiva. Los registros fueron proporcionados por un consultorio privado de la ciudad de Barquisimeto, Venezuela.

RESULTADOS

Mapa logístico

Nuestro primer estudio se realizó utilizando el mapa logístico,

$$x_{n+1} = r \times (1 - x_n) \quad (4)$$

ecuación que se aplica principalmente a problemas de crecimiento poblacional de especies animales o semejante índole. Siguiendo el esquema descrito anteriormente, hallamos el número de estados observables. Para ello hemos graficado el valor de la complejidad para diferentes cantidades de estados observables, figura 2. En este gráfico observamos como varía el valor de la complejidad a medida que N aumenta. Para un $N \rightarrow 1$, la complejidad es casi nula. La complejidad se incrementa a medida que N se hace mas grande, lo que quiere decir que la escala de

observación es muy pequeña. Sin embargo, para $N > 200$ la complejidad disminuye su nivel de variación y comienza a tener un crecimiento más regular. Esto nos indica que hemos alcanzado un valor de N capaz de resolver cerca del total del estados accesibles del sistema, por lo que el cálculo de la complejidad en este rango de parámetros será satisfactorio. El criterio básico para la elección de un N fijo consiste en verificar que la variación de la complejidad entre un N dado y el siguiente no sea muy grande. Para los cálculos siguientes, se ha considerado $N > 400$. Ahora bien, dado el número de estados, se puede obtener la distribución de probabilidades para este nivel de observación, y por tanto se puede calcular el valor de la complejidad. En la parte superior de la figura 3, se muestra un detalle del diagrama de bifurcación del mapa logístico. En esta figura, se observa la transición de caos a período tres y luego, la transición progresiva de período tres al caos. Esta transición es considerada uno de los ejemplos más conocidos en teoría de caos que muestra a un sistema que pasa de un estado periódico a un estado caótico, donde los niveles de complejidad deben ser mayores que 0. En la parte inferior de la figura 3 se muestra la variación de los niveles de complejidad para cada valor de parámetro y se observa efectivamente que la complejidad aumenta a medida que se incrementa la cantidad de períodos en el sistema, hasta que el sistema se vuelve caótico.

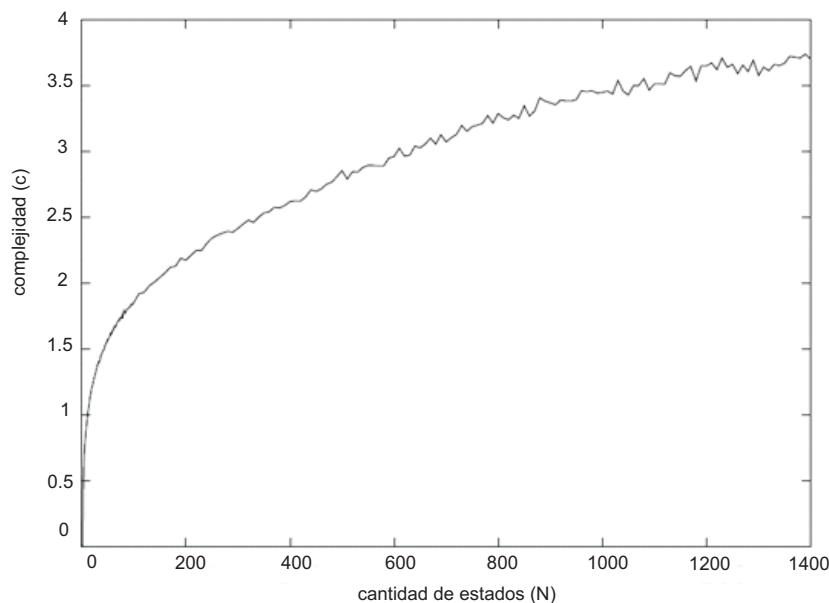


Figura 2. Variación de los niveles de complejidad con respecto a la cantidad de estados para una órbita caótica del mapa logístico.

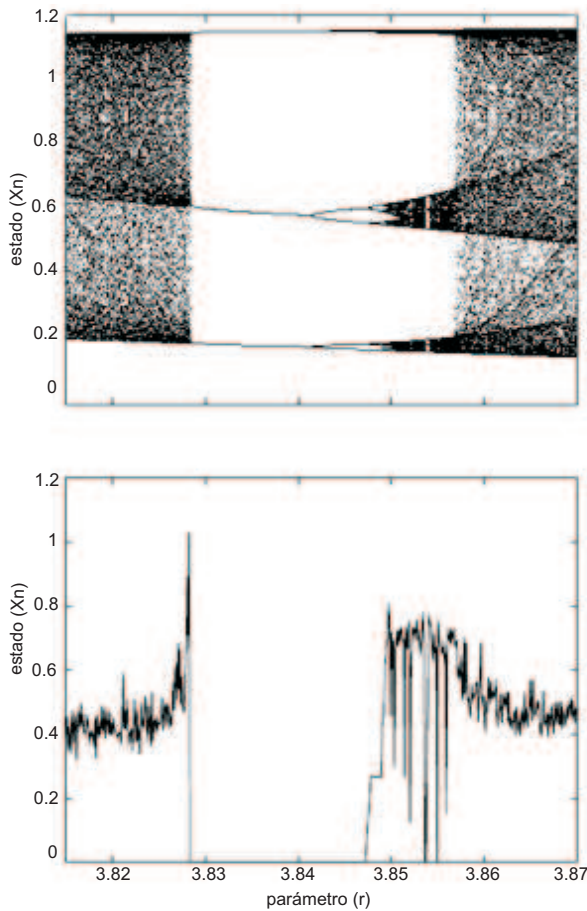


Figura 3. Arriba: Detalle del diagrama de bifurcación mostrado en la figura 1. Abajo: se muestran los niveles de complejidad para las diferentes regiones de la figura superior.

Señales Electroencefalografías

Los registros EEG han sido analizados siguiendo el esquema de ventanas de tiempo, de este modo los análisis se realizaron en un solo canal y no en el conjunto de canales del registro EEG. Los canales fueron elegidos con la ayuda de un médico especialista que determinó cuales canales poseen menor cantidad de ruido y por tanto cuales son susceptibles al estudio. Como se realizó anteriormente, es imprescindible obtener un valor apropiado para N . La variación de la complejidad con respecto a la cantidad de estados se muestra en la figura 4.

La línea gruesa de la figura 4 muestra la variación de la complejidad con respecto a N para un sujeto sano mientras que la línea fina muestra el mismo gráfico para un paciente con crisis epiléptica. Procediendo de la misma manera que en el mapa logístico, $N=600$

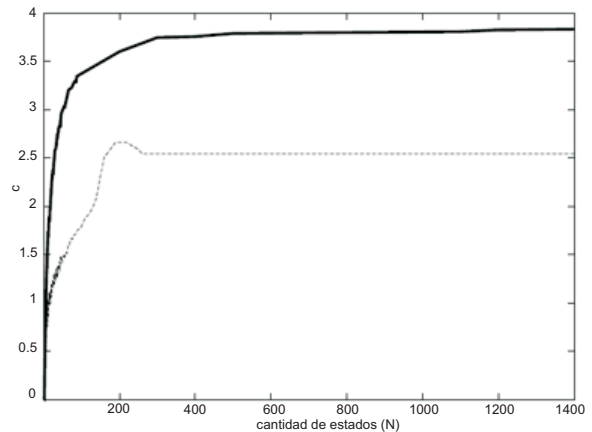


Figura 4. Complejidad con respecto al tamaño de ventana. La línea gruesa corresponde a un sujeto sano, mientras que la línea fina corresponde a un paciente epiléptico.

ha sido seleccionado para los cálculos siguientes. En el caso de los EEG, se desea conocer los diferentes niveles de complejidad en los diferentes estados del electroencefalograma, esto es, la variación de la complejidad en el tiempo. Para ello, debemos establecer un tamaño de ventana de tiempo o fracción del registro, en el cual realizaremos el cálculo de la complejidad. Para determinar el tamaño adecuado, mostramos a continuación la variación de la complejidad media con respecto al tamaño de la ventana (Figura 5).

En esta figura se observa un comportamiento similar al presentado en la figura 2 y 4, nuestro criterio de selección será por tanto similar al usado para la cantidad de estados. Nótese que al elegir un tamaño de ventana, para el caso de registros EEG, debe considerarse la no estacionariedad de la señal, sólo algunos segundos pueden ser considerados como estacionarios (Jiaw *et al.* 2005). Considerando el valor de la complejidad y el análisis de un médico especialista, el tamaño de la ventana se fijó en 1600, lo que equivale a unos 6,25 segundos de registro. Finalmente, la figura 6 muestra la evolución de la complejidad en un sujeto sano y en un paciente epiléptico que experimenta una crisis convulsiva en un instante indicado.

DISCUSIÓN DE RESULTADOS

Los análisis realizados para determinar el valor de N en cada sistema, i.e., mapa logístico y señales EEG, no ofrecen de manera directa el número exacto de estados accesibles del sistema, en su lugar, este tipo de análisis

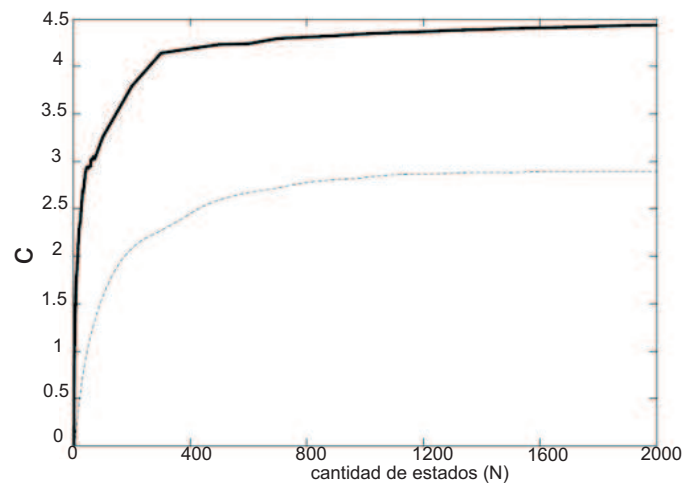


Figura 5. Complejidad con respecto al tamaño de ventana. Línea gruesa para un sujeto sano, línea fina para un paciente epiléptico

nos ofrece un rango de valores bastante amplio en el que se puede realizar la medida de la complejidad. La elección de un N se realiza

en el rango en el que la complejidad se mantiene casi constante, ya que esto asegura que una medida de complejidad en cualquiera de los valores del rango, se mantiene casi invariable. Por otro lado, el estudio sobre el tamaño de la ventana de tiempo realizado para el análisis de los registros EEG, muestra para ambos casos (sujeto normal y paciente epiléptico) un comportamiento similar al estudio de la cantidad de estados. Sin embargo, en este caso, es necesario tomar en cuenta que el tamaño de la ventana debe ser de por lo menos algunos cientos de datos o lo que equivale a algunos segundos, para evitar la no estacionariedad de la señal. Este tipo de análisis es más restrictivo que el caso de la cantidad de estados, ya que el valor está acotado por el lado inferior por la estabilización del valor de la complejidad y por la parte superior por la estacionariedad de la señal. En resumen los valores de N y tv presentan un rango de valores en los que no hay una gran variación de los resultados para un valor de N o tv y el siguiente. La medida de la complejidad realizada en las series de tiempo del mapa logístico, y mostradas para diferentes valores de parámetro (Figura 3), muestra la sensibilidad de este tipo de medida de complejidad. Para los diferentes valores de parámetro, el valor de la complejidad refleja el nivel de caoticidad del sistema. En la región caótica justo antes de iniciar el ciclo periódico, se observa que el valor de la complejidad es alto, una vez que el sistema entra en un ciclo periódico, el nivel de la complejidad se hace cero y luego, comienza a aumentar a medida que el sistema tiende al caos nuevamente. En el caso de los registros EEG,

podemos observar que los niveles de complejidad calculados para el sujeto sano son mayores que aquellos del paciente epiléptico, sugiriendo que la patología epiléptica es un estado de baja complejidad. Sin embargo, en el caso del paciente epiléptico, encontramos un incremento en el nivel complejidad en el momento del inicio de la crisis. Esto puede ser explicado por el aumento inesperado de la actividad eléctrica cerebral. A medida que la ventana de tiempo se desplaza por el registro realizando el cálculo de la complejidad, en el momento en que los primeros puntos de la crisis epiléptica se consideran para el cálculo, el sistema se hace muy impredecible y por tanto complejo, ya que la mayoría de los datos se encuentran en la zona sin epilepsia. A medida que más datos de la crisis son incorporados en el cálculo de la complejidad, el valor de complejidad disminuye. Una vez iniciado el episodio convulsivo, un decremento del nivel de complejidad es registrado, sugiriendo un estado menos caótico o más predecible instantes después del inicio de la crisis.

CONCLUSIONES

El método presentado en este artículo, para calcular la complejidad en series de tiempo, ha demostrado ser una herramienta útil y de fácil implementación, pues no requiere de altos recursos computacionales. Tanto las series temporales producidas por un modelo matemático como aquellas producidas por un sistema biológico —y que por tanto está expuesta al ruido y otros tipos de artefactos— son susceptibles a ser analizadas con la técnica presentada. Por otro lado, el conocimiento

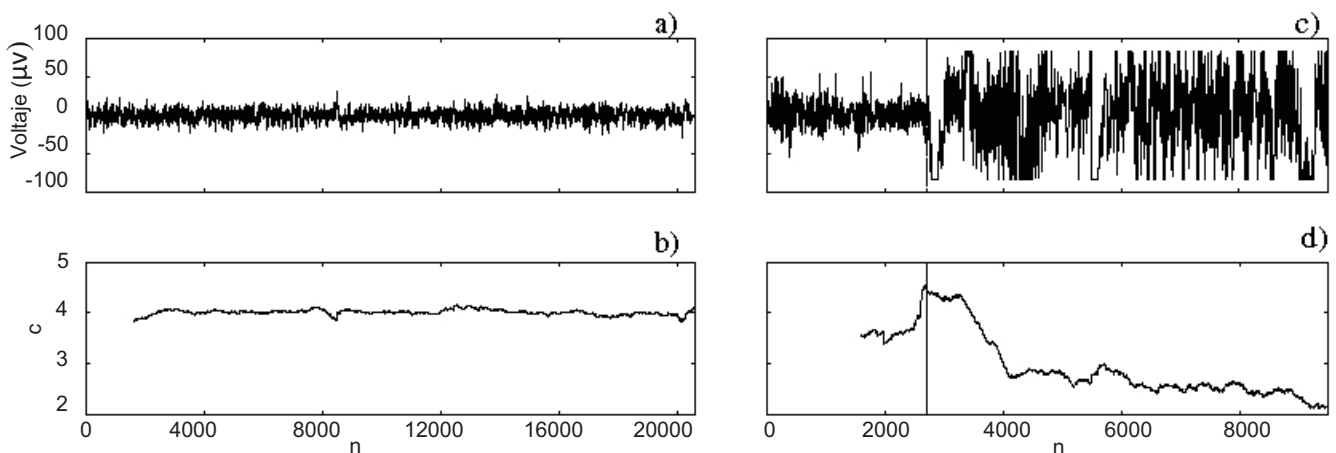


Figura 6. a) Señal EEG para un sujeto sano. b) Complejidad con respecto al tiempo correspondiente a (a). c) Señal EEG para un paciente epiléptico, antes y durante una crisis epiléptica cuyo inicio se indica por la línea vertical. d) Complejidad con respecto al tiempo correspondiente a (b)

de las leyes físicas que rigen la dinámica del sistema son importantes para definir los límites de los rangos de los parámetros necesarios para realizar el cálculo de la complejidad. Finalmente, los resultados del análisis de las señales EEG sugieren que la epilepsia es un estado fisiológico de menor complejidad en relación con el estado sano del cerebro.

AGRADECIMIENTOS

M. Escalona-Morán agradece al apoyo del Consejo de Desarrollo, Científico, Tecnológico, Humanístico y de las Artes, Universidad de los Andes, Mérida, mediante el proyecto I-121309-05-A.

REFERENCIAS BIBLIOGRÁFICAS

BABLOYANTZ, A. y DESTEXHE, A. Low dimension chaos in an instance of epilepsy. *Proc Natl Acad Sci USA*. pp. 83, 3513. 1986.
 BENNETT, C. En *Emerging syntheses in science*, ed. D. Pines, Addison-Wesley, MA. 1988.
 CALBET, X.; LÓPEZ-RUIZ, R. Tendency towards maximum complexity in an isolated non-equilibrium system. *Phys. Rev. E*, 63, 66116, 2001.
 CASDAGLI, M.; IASEMIDIS L.; SACKLELARES, J.; ROPER S.; GLIMORE R. y SAVIT, R. Characterizing nonlinearity in invasive EEG recordings from temporal lobe epilepsy. *Physica D* pp. 99, 381. 1996.

CHAITIN, G. *Information, randomness and incompleteness*. Singapur, World Scientific. 1990.
 GRASSBERGER P. Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.* pp. 25, 907, 1986.
 JIAW; KONG, N.; LI, F.; GAO, X.; GAO, S.; ZHANG, G.; WANG, Y.; YANG, F. An epileptic seizure prediction algorithm based on second-order complexity measure. *Physiol Meas.* pp. 26, 609. 2005.
 KOLMOGOROV, A. *Probl. Inform. Theory.* pp 1, 3. 1965.
 LEMPEL, A. y ZIV, J. *IEEE Trans. Inform. Theory.* P. 22. 1976.
 LÓPEZ-RUIZ, R.; MANCINI, H.; CALBET, X. *Phys. Lett. A. A Statistical Measure of Complexity.* *Phys. Lett. A.* pp. 209, 321, 1995.
 PALUS, M. Nonlinearity in human EEG: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biol Cybern.* pp. 75, 389. 1996.
 PERAKH, M. Defining complexity, obtenido de <http://www.talkreason.org/articles/complexity.pdf>, 2004.
 SHANNON, C. y WEAVER, W. *The mathematical theory of communication*, University of Illinois Press. 1949.
 SHATZ, C. *The Brain. Science.* pp. 214, 652-653, 1981.
 THELIER, J. y RAPP, P. Reexamination of the evidence for lowdimensional, non-linear structure in the human electroencephalogram. *Electroenceph Clin Neurophysiol* pp. 98, 213, 1996.