

Caracterización de la complejidad en señales electroencefalográficas

Miguel Escalona-Morán^{1}, Pedro García² y Mario G. Cosenza¹*

¹Centro de Física Fundamental, Área de Caos y Sistemas Complejos. Universidad de los Andes, Mérida, Venezuela. ²Facultad de Ingeniería. Universidad Central de Venezuela, Caracas, Venezuela.

Recibido. 22-10-06 Aceptado: 09-10-07

Resumen

Se presenta un algoritmo para calcular la complejidad de un sistema con dimensión alta a partir de datos experimentales, basado en la definición de complejidad estadística introducida por López-Ruiz, Mancini y Calbet (16). Mediante este algoritmo, se calcula la complejidad de señales electroencefalográficas provenientes de diez sujetos sanos y treinta pacientes epilépticos. La dimensionalidad de las señales se reduce mediante el método de Análisis de Componentes Principales. Los valores de complejidad estadística así obtenidos permiten caracterizar los dos grupos de individuos. Los resultados muestran que la complejidad de los estados colectivos cerebrales asociados con la patología epiléptica es menor que aquella correspondiente a sujetos sanos.

Palabras clave: electroencefalograma; epilepsia; sistemas complejos.

Characterization of complexity in electroencephalographic signals

Abstract

Based on the definition of statistical complexity introduced by López-Ruiz, Mancini y Calbet (16), we present an algorithm to calculate the complexity of a system from experimental data. By using this algorithm, the complexity of electroencephalographic signals is calculated. The data base consists of 10 healthy subjects and 30 epileptic patients. The Principal Component Analysis method has been employed to reduce the dimensionality of the signals. The values of the statistical complexity obtained in this way allow to characterize the two groups of individuals. The results show that the complexity of the collective brain states associated to the epileptic pathology is lower than that corresponding to healthy subjects.

Key words: Complex systems; electroencephalogram; epilepsy.

Introducción

El estudio de las propiedades macroscópicas emergentes, tanto espaciales como temporales, en sistemas de elementos dinámicos no lineales interactivos es objeto

de mucha atención en la literatura científica actual. Fenómenos como formación espontánea de estructuras, organización, patrones espaciotemporales, sincronización y oscilaciones colectivas, ondas espirales, segre-

* Autor para la correspondencia. E-mail: angele@ula.ve.

gación y diferenciación, formación y crecimiento de dominios y de interfases, etc., son ejemplos de procesos de auto-organización que ocurren en diversos contextos, tales como sistemas físicos, químicos, biológicos, sociales, económicos, etc. Estos efectos colectivos emergentes son el resultado de la interacción y cooperación entre los elementos constituyentes de un sistema. Avances teóricos, computacionales y experimentales en este campo están contribuyendo aceleradamente al surgimiento de una descripción unificada de las propiedades de estos sistemas. 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 sistemas complejos (1-4) son sistemas de elementos interactivos cuyas propiedades colectivas no son susceptibles de ser derivadas trivialmente a partir del conocimiento o análisis reduccionista de las propiedades de sus elementos constituyentes. Un sistema complejo se encuentra a medio camino entre lo desordenado (un gas, por ejemplo) y lo ordenado (un cristal). Un sistema totalmente aleatorio es estadísticamente invariante en el tiempo y/o el espacio y no produce estructuras. Por otro lado, un sistema regular, ordenado en el espacio y/o el tiempo, es fácilmente reproducible a partir del conocimiento de una parte de ese sistema. En un sistema complejo existen elementos de desorden que hacen difícil predecir la estructura global a partir de fragmentos de esta estructura. Al mismo tiempo, subsiste un orden subyacente dentro de esta estructura. Los sistemas complejos se caracterizan por exhibir variabilidad y diversidad.

En particular, existe actualmente mucho interés en el estudio de sistemas fisiológicos, especialmente el cerebro, desde el punto de vista de sistemas complejos. Numerosas investigaciones intentan describir las señales cerebrales, tales como los electroencefalogramas (EEG), como propiedades colectivas de un sistema dinámico no lineal complejo (5-7). El cerebro humano es

frecuentemente descrito como el más complejo de los sistemas conocidos. Consiste en una red altamente interconectada de neuronas. El comportamiento dinámico de una neurona individual es simple. Sin embargo, a partir de la señal de una neurona, o aún de un grupo de neuronas, no pueden deducirse las funciones y propiedades colectivas sofisticadas exhibidas por el cerebro.

Varias definiciones han sido propuestas para cuantificar el grado de complejidad de un sistema, dependiendo de los aspectos colectivos o funcionales que se intentan caracterizar. En general, estas definiciones de complejidad se pueden clasificar en tres categorías:

1. *Complejidad algorítmica*; de uso frecuente en las ciencias computacionales, se refiere a la caracterización de la complejidad de un conjunto de datos como el programa más corto (medido en bits) que puede reproducir esos datos. En esta categoría se encuentran la complejidad de Kolmogorov-Chaitin (8,9), la complejidad de Lempel-Ziv (10) y la profundidad lógica de Bennett (11).

2. *Complejidad predictiva*; corresponde a la cantidad de información requerida para predecir el comportamiento de un sistema. Esta categoría se basa en técnicas de análisis no lineal de series de tiempo para describir el flujo de información entre las partes del sistema. Ejemplos de este tipo son la medida efectiva de la complejidad de Grassberger (12), la complejidad basada en diversidad (13), la profundidad termodinámica (14) y la información mutua (15), entre otras.

3. *Complejidad estructural*; este punto de vista se enfoca en la emergencia de estructuras macroscópicas en sistemas complejos. A esta categoría pertenece la medida de complejidad estadística introducida por López-Ruiz, Mancini y Calbet (LMC) (16).

Cabe señalar que la mayoría de estas caracterizaciones de la complejidad se han aplicado principalmente en modelos teóricos o computacionales, ya que su implementa-

ción en la práctica resulta complicada. Por otro lado, la medida de complejidad estadística LMC está basada en una noción intuitiva que permite discernir diversas situaciones reconocidas como complejas.

En este artículo presentamos un algoritmo general y eficiente para calcular la medida de complejidad estadística LMC a partir de series de tiempo experimentales. Específicamente, calculamos la complejidad LMC en señales EEG en grupos de sujetos sanos y de pacientes diagnosticados con epilepsia. Esta medida permite caracterizar ciertos aspectos del funcionamiento del cerebro humano como un sistema complejo. Se muestra que la complejidad LMC permite establecer un criterio cuantitativo para diferenciar entre los dos grupos de individuos. Nuestros resultados revelan que la patología epiléptica está asociada a estados cerebrales con menor complejidad estadística que aquellos estados correspondientes a sujetos sanos.

A continuación, presentamos una breve revisión de la definición de la complejidad estadística LMC. Luego, se describe la base de datos de los EEG usados en este trabajo. El proceso de reducción de la dimensionalidad de las señales EEG empleando el método de Análisis de Componentes Principales es explicado después. Más adelante, los resultados son mostrados y discutidos. Finalmente, se presentan las conclusiones del presente estudio.

Complejidad LMC

Asumamos que a una cierta escala de observación, o nivel de descripción, un sistema posee 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 (17) 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 contenida en un cristal ordenado es $H_{\text{cristal}} \approx 0$, mientras que un gas aislado térmicamente tiene una distribución equiprobable, $p_i = 1/N$, y entonces $H_{\text{gas}} \approx -k \log N$; 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, Mancini y Calbet (16) 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 (18).

La complejidad de un sistema resulta de la competencia entre estas dos cantidades. López-Ruiz, Mancini y Calbet (LMC) definen la complejidad estadística de un sistema como el producto de la cantidad de información almacenada en el sistema y su desequilibrio (16),

$$C = H \cdot D = - \left(k \sum_{i=1}^N p_i \log p_i \right) \times \left(\sum_{i=1}^N \left(p_i - \frac{1}{N} \right)^2 \right). \quad [3]$$

Esta definición concuerda con las propiedades que, de manera intuitiva, debería poseer una 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 interpola-

ción 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 (19).

Nótese que el valor de la complejidad LMC de un sistema depende de la escala de observación, la cual determina el número de estados accesibles 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.

Base de datos

El registro de las señales EEG para cada sujeto se realizó en 19 canales a partir de electrodos adheridos con pasta conductora a la superficie de la cabeza en los sitios establecidos por el sistema internacional 10-20. La diferencia de potencial se midió con respecto a una referencia constituida por ambas orejas cortocircuitadas. La señal se digitalizó con una frecuencia de muestreo de 256 Hz y conversión A/D de 12 bits, filtrada digitalmente entre 0,5 Hz y 30 Hz. Para el análisis, se seleccionaron registros de tiempo variable de actividad en reposo de diez sujetos sanos, diez pacientes epilépticos bajo tratamiento médico, dieciocho pacientes epilépticos sin tratamiento médico y dos pacientes epilépticos durante crisis convulsivas. Cada registro electroencefalográfico corresponde a un sujeto distinto. Las ba-

ses de datos utilizados en este trabajo provienen del Hospital San Juan de Dios, en Mérida, y del Hospital José María Pineda, en Barquisimeto, Venezuela.

Procesamiento de las señales EEG

Nos enfocaremos en el cálculo de la complejidad LMC correspondiente a la emergencia de estructuras o patrones macroscópicos en las señales EEG. Cada señal EEG corresponde a una serie temporal que se puede describir mediante un vector x_t^c , donde $c = 1, 2, \dots, 19$ identifica el canal del EEG y t corresponde al tiempo discreto del muestreo del canal que puede contener hasta 20.000 puntos. Se calcula la variabilidad de la señal EEG, definida como $s_t^c = x_t^c - \bar{x}^c$, donde $\bar{x}^c = \frac{1}{T} \sum_{t=1}^T x_t^c$ es el valor medio temporal del canal c .

Luego, el vector asociado a cada señal tiene una dimensión igual a diecinueve. El cálculo de la complejidad LMC requiere realizar una partición de un espacio de diecinueve dimensiones para establecer la distribución de probabilidad correspondiente a unos veinte mil puntos por cada señal EEG, lo cual evidentemente se vuelve impráctico. Para implementar un cálculo eficiente de la complejidad LMC, hemos empleado el método conocido como *Principal Component Analysis* (PCA) (20), el cual permite reducir la dimensionalidad efectiva de un conjunto de datos y se utiliza en aplicaciones de compresión de datos y de procesamiento de imágenes. El método PCA consiste en una transformación de los datos (mediante una función llamada *kernel*) a un nuevo sistema de coordenadas tal que la mayor varianza de cualquier proyección de los datos yace en la primera coordenada (denominada la componente principal), y así sucesivamente. El procedimiento seguido para obtener la complejidad de cada señal EEG es el siguiente:

1. Se realiza una transformación del vector de variabilidad s_i^c a un nuevo espacio de funciones por medio de un kernel PCA no lineal definido abajo.

2. Se calculan los autovalores y autovectores de la matriz de transformación de tamaño 19×19 y se seleccionan, para el análisis posterior, aquellos autovectores en cuya dirección se encuentra la mayor varianza de los datos. El número de estos autovalores corresponde a la dimensión reducida del vector s_i^c transformado.

3. Se establece una escala apropiada para medir la distribución de probabilidades del vector transformado.

4. Se calcula el valor de complejidad LMC de acuerdo con la ecuación [3].

Como función de transformación para el método PCA no lineal hemos utilizado el kernel gaussiano (21)

$$K(c, c') = \exp\left(-\frac{|s_i^c - s_i^{c'}|}{2\sigma^2}\right), \quad [4]$$

donde s_i^c y $s_i^{c'}$ son componentes del vector de variabilidad del EEG y σ es un parámetro que define el ancho de las gaussianas que constituyen las bases del espacio de funciones. Para elegir un σ adecuado, se encuentra un valor para el cual la matriz de transformación constituida por las combinaciones (c, c') no sea singular. Para nuestra base de datos, esta condición se cumple para $\sigma \geq 240$. En el resto de este trabajo hemos utilizado $\sigma = 248$. Un valor de σ mayor nos asegura la invertibilidad de la matriz; sin embargo, si $\sigma \rightarrow \infty$, no se observarán correlaciones entre los datos, ya que no existe criterio para discernir entre un vector y otro.

En la Figura 1 se muestra el espectro de autovalores de la matriz de transformación correspondientes a un sujeto sano y a un paciente epiléptico representativos de la base de datos. En esta Figura observamos que en ambos casos el número de autovalores de mayor magnitud es 3; los demás autovalores tienen

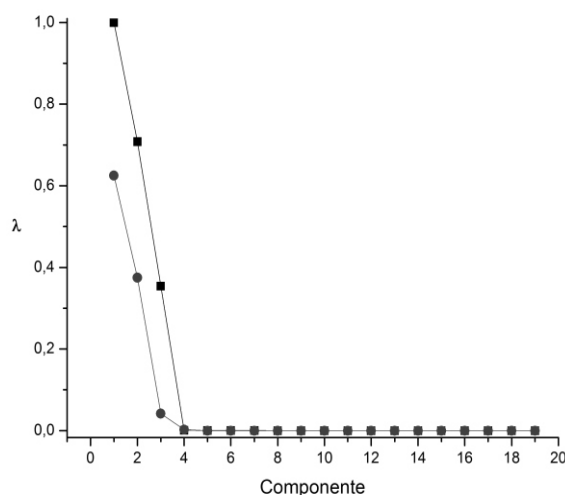


Figura 1. Espectro de autovalores de la matriz de transformación asociada al método PCA con kernel gaussiano, para un sujeto sano (cuadrados) y para un paciente epiléptico (círculos) pertenecientes a la base de datos de las señales

una magnitud despreciable. Luego, la mayor varianza de los datos se encuentra en las direcciones de los autovectores o componentes principales asociados a estos 3 autovalores; esto es, los vectores que describen los EEG pueden ser representados en una base que define un espacio tridimensional sin pérdida significativa de información. Mediante este método de reducción efectiva de la dimensionalidad de los vectores EEG, es posible implementar de manera práctica y eficiente el cálculo de la complejidad estadística LMC para nuestra base de datos.

Resultados y Discusión

En la Figura 2 se muestran los valores de la complejidad C (ecuación [3]) para los 40 EEG disponibles en la base de datos. La región I de la Figura 2 corresponde al grupo de sujetos sanos; la región II contiene a los pacientes con epilepsia bajo tratamiento médico; en la región III se encuentran los pacientes con epilepsia sin tratamiento médico; y

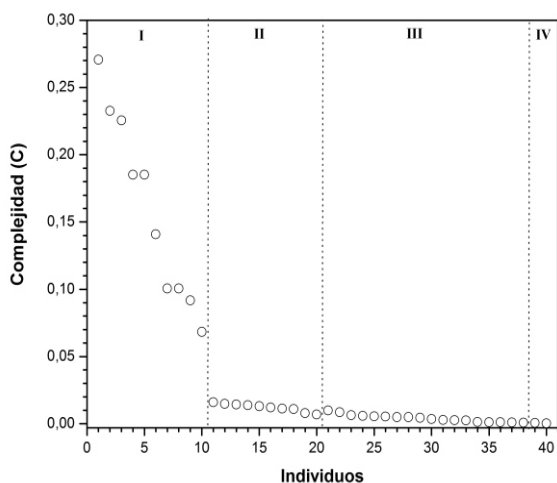


Figura 2. Distribución de valores de la complejidad estadística para las 40 señales EEG disponibles en la base de datos. Región I: grupo de sujetos sanos; región II: pacientes con epilepsia bajo tratamiento médico; región III: pacientes con epilepsia sin tratamiento médico; región IV: pacientes con epilepsia durante una crisis convulsiva. El orden de los individuos dentro de cada región no es relevante.

en la región IV se hallan los pacientes con epilepsia durante una crisis convulsiva.

Los resultados de la Figura 2 indican que la complejidad estadística de los estados colectivos cerebrales asociados con la patología epiléptica (regiones II, III y IV) es menor que aquella correspondiente a sujetos sanos (región I). Otros estudios de señales EEG, desde el punto de vista de la dinámica no lineal, han encontrado que la epilepsia está asociada a estados de mayor sincronización de las neuronas en el cerebro (6, 7). La mayor sincronización que tiene lugar en la patología epiléptica significa que, en este caso, el comportamiento colectivo del cerebro es más predecible a partir del conocimiento del comportamiento local y, por

tanto, esto se refleja en un grado de complejidad menor en relación con el cerebro de un sujeto sano.

Podemos observar que los valores de complejidad de los pacientes epilépticos con tratamiento médico (región II) es, en promedio, ligeramente mayor que la complejidad de los pacientes sin tratamiento (región III), lo que sugiere una posible influencia de los fármacos en el aumento de la complejidad cerebral, sin llegar a los valores de C asociados con sujetos sanos. Nótese que en los dos episodios de crisis convulsivas (región IV) los valores de complejidad decrecen aún más, lo cual implica que el comportamiento colectivo del cerebro sería aún más fácil de predecir en esos casos.

Como se ha dicho, el valor de la complejidad LMC depende de la escala de observación. En el caso de las señales EEG, esta escala viene representada por el número de cajas o particiones N realizadas para empotrar los datos en el espacio de los 3 ejes principales obtenidos con el método PCA. La Figura 3 muestra el valor de la complejidad en función de $N^{1/3}$, correspondiente a diversas escalas de observación. En la Figura 3 observamos que el valor de C aumenta a medida que se incrementa el tamaño de la partición y, por lo tanto, de la cantidad de estados accesibles. Nótese que el valor de la complejidad es prácticamente invariante para $N^{1/3} \geq 11$. Los cálculos correspondientes a las Figuras 1 y 2 fueron realizados con $N^{1/3} = 22$; es decir, con $N = 10648$ estados accesibles.

Conclusiones

En este artículo hemos mostrado que la medida de complejidad LMC puede ser calculada en forma eficiente a partir de datos experimentales, tales como señales EEG. Esta medida permite caracterizar el cerebro humano como un sistema dinámico complejo. Hemos calculado la complejidad LMC en señales EEG provenientes de grupos de su-

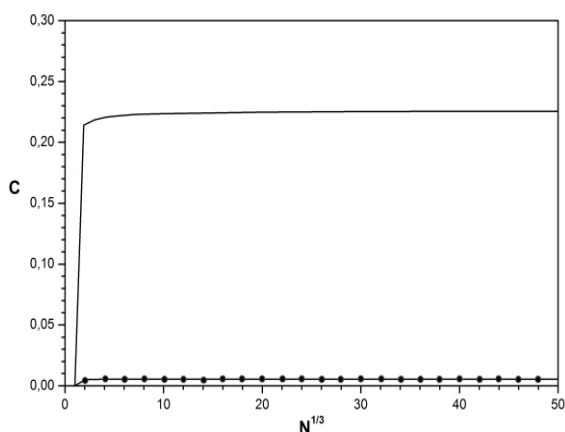


Figura 3. Complejidad en función del tamaño de la partición $N^{1/3}$ correspondiente a diversas escalas de observación para una señal EEG de un sujeto sano (línea continua) y para un paciente epiléptico (línea con círculos).

jetos sanos y de pacientes diagnosticados con epilepsia. La medida de la complejidad LMC permite diferenciar entre los dos grupos de individuos. Nuestros resultados han revelado que la patología epiléptica está asociada, en general, a estados cerebrales con menor complejidad estadística que aquellos estados correspondientes a sujetos sanos. Igualmente, se ha encontrado que existen diferencias perceptibles en la complejidad del comportamiento colectivo del cerebro en los tres subgrupos de pacientes epilépticos; en particular, los EEG correspondientes a las crisis convulsivas exhiben el menor grado de complejidad de todos los grupos de individuos.

Los resultados del presente trabajo sugieren que el empleo del concepto de sistema complejo y de herramientas para la caracterización de la complejidad del cerebro a partir de señales EEG pueden contribuir a una mejor comprensión y diagnóstico de algunas patologías cerebrales. El algoritmo general presentado aquí puede ser útil para caracterizar la complejidad estadística de otras señales fisiológicas, como electrocardiogramas, y de

diversos sistemas dinámicos mediante el análisis de datos experimentales, como registros sísmicos, series económicas, etc.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Consejo de Desarrollo Científico, Humanístico y Tecnológico de la Universidad de los Andes, mediante el proyecto N° C-1396-06-05-B, y por el Fondo Nacional de Ciencia, Tecnología e Innovación (Fonacit), a través del proyecto N° F-2002000426.

Referencias Bibliográficas

1. KANEKO K., TSUDA I. **Complex systems: chaos and beyond**. Springer, 2000.
2. BOCCARA N. **Modeling complex systems**. Springer, 2004.
3. MIKHAILOV A., CALENBUHR V. **From cells to societies: models of complex behavior**. Springer, 2002.
4. SCHUSTER H. G. **Complex adaptive systems**. Scator-Verlag, 2002.
5. RAPP P. E., BASHORE T., MARTINERIE J., ALBANO A., ZIMMERMAN I., MESS A. **Brain Topography** 2: 99-118, 1989.
6. CASDAGLI M.C., IASEMIDIS L.D., SAVIT R.S., GILMORE R.L., ROPER S.N., SACKLELARES J.C. **Electroencephalogr Clin Neurophysiol** 1: 98-102, 1997.
7. ESCALONA-MORÁN, M., GUILLÉN P., COSENZA, M.G., COUTIN, P. **Ciencia** 13: 505-510, 2005.
8. KOLMOGOROV A.N. **Probl Inform Theory** 1: 3-12, 1965.
9. CHAITIN G.J. **Information, randomness and incompleteness**. World Scientific, Singapur (Singapur), 1990.
10. LEMPELA A., ZIV J. **IEEE Trans. Inform. Theory** 22: 75-81, 1976.
11. BENNET C.H. **Emerging syntheses in sciences**. D. Pines (Editor), Adison-Wesley, Reading (Inglaterra), 1988.
12. GRASSBERGER P. **Int J Theor Phys** 25: 907-915, 1986.

13. HUBERMAN B.A., HOGG T. *Physica D* 22: 376-384, 1986.
14. LLOYD S., PAGELS, H. *Ann Phys* 188: 186-188, 1988.
15. ARNHOLD J., GRASSBERGER P., LEHNERTZ K., ELGERT C.E. *Physica D* 134: 419-425, 1999.
16. LÓPEZ-RUIZ R., MANCINI H.L., CALBET X. *Phys Lett A* 209: 321-326, 1995.
17. SHANNON C.E., WEAVER W. *The mathematical theory of communication*. University of Illinois Press, Champaign (Estados Unidos), 1949.
18. CALBET X., LÓPEZ-RUIZ R. *Phys. Rev. E* 63: 066116, 2002.
19. LÓPEZ-RUIZ R. *Int J Bifurcations & Chaos* 11: 2669-2673, 2001.
20. SCHOLKOPF B., BURGESS, C.J.C., SMOLA A.J. (Editores) *Advances in kernel methods-Support vector learning*. MIT Press, Cambridge (Estados Unidos), 1999.
21. CRISTIANINI N., SHAWE-TAYLOR J. *An introduction to support vector machines and other kernel-based methods*. Cambridge University Press, Cambridge (Estados Unidos), 2001.