



Revista Electrónica de Metodología Aplicada
2001, Vol. 6 nº 1, pp. 1-12

EL GRAFICO DE RECURRENCIA EN EL ANALISIS DE SERIES TEMPORALES: EL PROGRAMA VRA (Visual Recurrence Analysis).

Ana María López Jiménez*
Carlos Camacho Martínez Vara de Rey*
Antonio Ramón García Torres**
José Enrique Romero García***

***Facultad de Psicología (Universidad de Sevilla)**

****I.E.S. (Los Viveros, Sevilla)**

***** Departamento de Economía Aplicada (Universidad de Sevilla)**
e-mail:analopez@cica.es

ABSTRACT.

In the last years different procedures have been proposed that allow to differentiate stochastic variability of chaotic variability. Among these procedures it is the recurrence plot. In this work we describe the utilities implemented in the package VRA to build recurrence plots and to quantify their most relevant parameters.

Key words: recurrence plot, determinism, entropy, nonlinear dynamics, chaos.

Agradecemos a D. Antonio Córdoba Zurita (Catedrático de Física de la Universidad de Sevilla) los acertados comentarios a este manuscrito.

1. Introducción.

El descubrimiento del comportamiento caótico en sistemas dinámicos deterministas ha modificado aspectos filosóficos del paradigma científico dominante y ha abierto perspectivas analíticas nuevas para los diseños de series de tiempo (Barnett y Choi, 1989; Berliner, 1992; Casdagli, 1991; Casdagli et al., 1991; Gerr y Allen, 1993; López, 1998; López, Camacho y García, 2001; McCaffrey et al., 1992; Nychka et al., 1992; Sayers, 1991; Takens, 1993).

Los progresos, en la década de los 80, en el análisis de series de tiempo desde la perspectiva de la Teoría Cualitativa de los Sistemas Dinámicos han dado lugar a una serie de índices analíticos y procedimientos gráficos que permiten, en teoría, determinar si una serie de tiempo aparentemente aleatoria puede o no corresponder a comportamiento caótico generado por un sistema de ecuaciones deterministas no lineales. Como subproducto es posible determinar el número de variables que pondría en juego este conjunto de ecuaciones desconocidas, clasificar sistemas dentro de clases universales (lineales-no lineales, estocásticos deterministas) y relacionar cambios en los cuantificadores del comportamiento con cambios en la conducta dinámica del sistema (bifurcaciones) (López, 1998; Montero y Morán, 1992; Sugihara y May, 1990).

Dimensión de correlación, exponentes de Lyapunov y entropía de Shannon se encuentran entre los procedimientos más utilizados para el análisis de series temporales en el contexto de la Teoría Cualitativa de los Sistemas Dinámicos (Abarbanel, 1993; Casdagli, 1997; Eckmann, Kamphorst y Ruelle, 1987; López, 1998; López, Camacho y García, 2001). Todos estos procedimientos requieren, no obstante, para su correcta aplicación e interpretación que las series sean estacionarias. Dado que muchas series experimentales no lo son es necesario desarrollar otras técnicas que, manteniendo su sensibilidad a las diferencias entre comportamiento aleatorio y caótico, permitan su análisis. El gráfico de recurrencia propuesto por Eckmann, Kamphorst y Ruelle en 1987 es adecuado para todo tipo de series con la única condición de que los datos estén igualmente espaciados. Series de tiempo estacionarias y no estacionarias, lineales y no lineales pueden ser analizadas con el gráfico de recurrencia (Casdagli, 1997; Trulla, Giuliani, Zbilut y Webber, 1996).

En la primera parte de este trabajo vamos a definir y describir las características básicas del gráfico de recurrencia que permiten distinguir entre variabilidad aleatoria y caótica y en la segunda parte describimos las utilidades implementadas en el programa VRA (Visual Recurrence Analysis) para construir gráficos de recurrencia y cuantificar sus parámetros más relevantes.

2.- Gráficos de recurrencia.

Si consideramos una serie de tiempo unidimensional, X_1, X_2, \dots, X_N , generada de un sistema dinámico determinista de dimensión d , Mañé, (1981), Packard et al. (1980) y Takens (1981) demostraron que es posible, a partir de esta serie correspondiente a una cualquiera de

las variables del sistema, obtener información del sistema completo. Demostraron que se puede generar un espacio multidimensional denominado espacio de estados artificial, reconstruido o pseudoespacio de fases ("embedding space") de dimensión $d_e \geq 2d + 1$, de manera que las trayectorias en el espacio reconstruido tienen las mismas propiedades geométricas y dinámicas que las trayectorias en el espacio multidimensional del sistema dinámico completo.

Para la construcción del espacio de estados artificial se parte de la serie temporal y se agrupan los valores de la misma formando "vectores" cuyas coordenadas son los valores de la serie y sus consiguientes retardos. El número de coordenadas es igual a la dimensión del espacio artificial (d_e) y el número de vectores que se pueden construir es

$$M = N - d_e$$

Un vector de retardos en una dimensión de vendría dado por $(X_i, X_{i+T}, \dots, X_{i+(d_e-1)T})$, donde el retardo T puede ser cualquier valor. Si hacemos $T = 2$ y $d_e = 4$ los vectores de retardo se obtendrían agrupando los valores de la serie tal y como se muestra en la Figura 1.

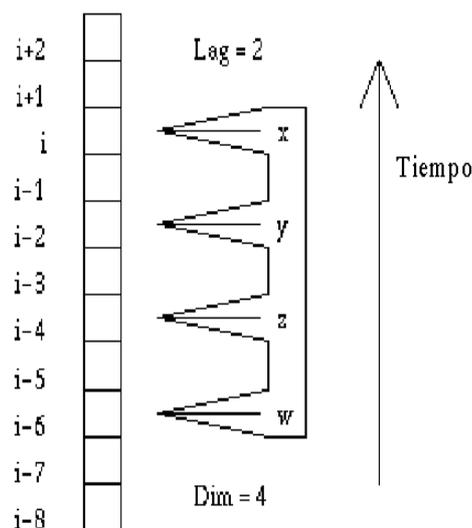


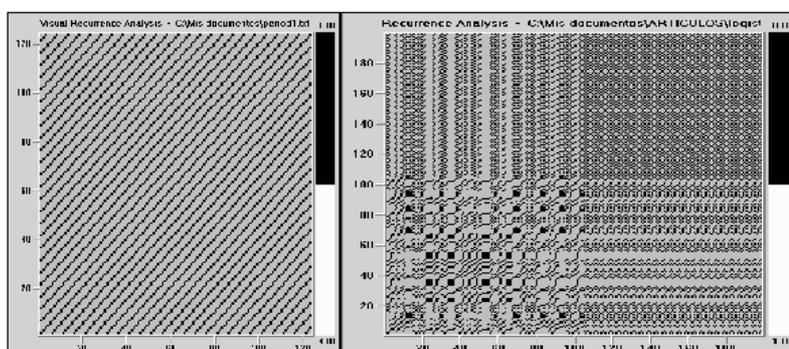
Figura 1. Método de retardos para la construcción de vectores.

Después de elegir una dimensión de reconstrucción (d_e) y obtener los correspondientes vectores, el gráfico de recurrencia es un conjunto de puntos en un cuadrado de dimensión $M \times M$, donde M , los ejes o los lados del cuadrado, representan la secuencia cronológica de los vectores en el espacio reconstruido. Para obtener el gráfico de recurrencia se coloca un punto en la intersección de las coordenadas (i, j) si el vector $x(i)$ está próximo a $x(j)$. El criterio de proximidad es una distancia pequeña previamente seleccionada (Eckmann, Kamphorst y Ruelle, 1987; Casdagli, 1997). Dado que i y j son realmente tiempos el gráfico de recurrencia da información del alcance temporal de la correlación entre los valores de la serie.

La inspección visual del gráfico de recurrencia mostrará M puntos en la bisectriz del cuadrado. Además, pueden aparecer segmentos de línea más cortos paralelos a la bisectriz y correspondientes a vectores, cronológicamente secuenciados, que cumplen con el criterio de proximidad establecido. Teóricamente, estos segmentos no deben aparecer si el mecanismo generador de la serie es aleatorio y aparecerán en caso de sistemas deterministas. Si la serie fuera periódica en el gráfico de recurrencia aparecerían segmentos de recurrencia paralelos a la bisectriz. Con fines exclusivamente didácticos si quisiéramos construir el gráfico de recurrencia a partir de una serie periódica ($p = 5$) en un espacio de dimensión 1 y con un retardo también de 1, calcularíamos, en primer lugar, las distancias entre los valores de la serie. Por ejemplo, si la serie viene dada por 24 secuencias de los valores: 1, 2, 3, 4 y 5 la matriz de distancias entre dichos valores es:

	1	2	3	4	5
1	0	1	2	3	4
2		0	1	2	3
3			0	1	2
4				0	1
5					0

En segundo lugar, y si consideramos que dos vectores están próximos cuando la distancia entre ellos es $d(x(i), x(j)) \leq 1$, dibujaríamos un punto en la intersección de los pares de vectores que cumplen con el criterio de proximidad. El gráfico a) de la Figura 2 es el gráfico de recurrencia de la serie formada con las 24 secuencias de los valores 1, 2, 3, 4, 5. En dicho gráfico son claramente apreciables los segmentos paralelos a la bisectriz que indican periodicidad.



a)

b)

Figura 2. a) Gráfico de recurrencia de una serie periódica. b) Gráfico de recurrencia de la transición de un estado caótico a uno periódico ($p=8$).

El gráfico de recurrencia es un procedimiento cualitativo para diferenciar variabilidad caótica de estocástica o para detectar cambios de estado en la evolución de un sistema (ver gráfico b) de la Figura 2). Desarrollos posteriores han permitido cuantificar la cantidad de recurrencia presente en el gráfico. Zbilut y Webber en 1992 proponen el RQA (Recurrence Quantification Analysis) definido por los siguientes índices:

a) Porcentaje de puntos recurrentes (%REC): es una medida global de recurrencia. Dado que el gráfico de recurrencia es simétrico con respecto a la bisectriz del cuadrado $M \times M$, el %REC es

$$\% REC = \frac{NRECURS}{\frac{M(M-1)}{2}} \cdot 100 \quad (1)$$

donde, NRECURS es el número total de puntos recurrentes en el triángulo superior del gráfico sin contar los puntos de la bisectriz. El denominador de (1) es el total de puntos en el triángulo superior del gráfico eliminados los de la bisectriz.

b) Porcentaje de determinismo (%DET) viene dado por

$$\% DET = \frac{DRECURS}{NRECURS} \cdot 100 \quad (2)$$

donde, DRECURS es el número de puntos que forman parte de segmentos de línea paralelos a la bisectriz del cuadrado. Un segmento de línea se define como dos o más puntos adyacentes.

c) Entropía de Shannon o información de Shannon(ENT): es una medida de la información media contenida en el gráfico de recurrencia y se obtiene del histograma de la longitud de los segmentos de puntos recurrentes paralelos a la bisectriz. La entropía de Shannon se calcula como

$$ENT = - \sum_{i=1}^n p_i \log p_i \quad (3)$$

donde, p_i es la frecuencia relativa de la longitud de los segmentos recurrentes. Si la serie es aleatoria está medida es 0.

d) Una medida de sensibilidad a las condiciones iniciales viene dada por la longitud del segmento recurrente más largo (MAXLINE). MAXLINE es inversamente proporcional al máximo exponente de Lyapunov para sistemas caóticos.

e) Una medida de estacionaridad de la serie (TREND). Se obtiene ajustando una recta de mínimos cuadrados al porcentaje de recurrencia en función del desplazamiento de la bisectriz. Si la pendiente es un valor próximo a cero es una indicación de estacionaridad.

Todos estos índices están implementados en el programa VRA que vamos a describir a continuación.

3.- El programa VRA (Visual Recurrence Analysis v4.0).

El VRA es un programa escrito por Eugene Kononov basándose en la definición de gráfico de recurrencia de Eckmann, Kamphorst y Ruelle (1987) y en el RQA de Zbilut y Webber (1992). Es un programa que se distribuye gratuitamente desde el sitio <http://pw1.netcom.com/~eugenek/download.html>. Se ejecuta en el entorno Windows (Windows 95, Windows 98 o Windows NT) y necesita al menos 8 Mb de RAM y 1 Mb de espacio libre en el disco duro.

Para instalar el programa se descomprime el archivo vra4r0.zip y se ejecuta el archivo setup.exe. Una vez instalado, obtener el gráfico de recurrencia de una serie temporal requiere abrir mediante la opción "Open Signal" del menú "File" el archivo que contiene la serie. El programa, siempre que la serie esté escrita en una columna con un mínimo de 100 datos y un máximo de 1.000.000, soporta los siguientes formatos: ASCII, .csv, .prn, WAV y .xls. La figura 3 muestra el gráfico de recurrencia correspondiente a una serie de 1000 datos generados del sistema de Lorenz. El archivo que contiene la serie se denomina lorenz.dat y es uno de los 15 archivos de datos que se instalan con el programa.

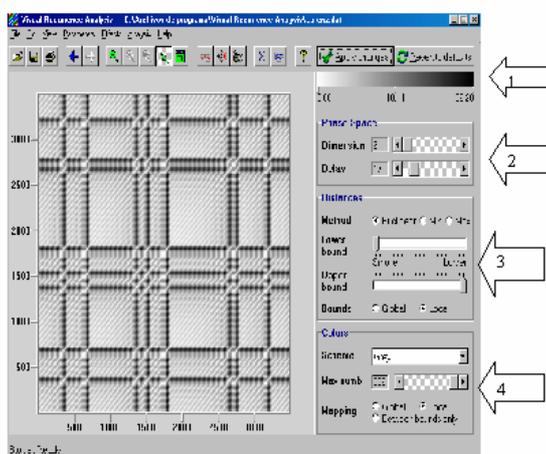


Figura 3. Gráfico de recurrencia para la serie de Lorenz.

Por defecto el programa VRA hace corresponder puntos de colores distintos dependiendo de la distancia entre los vectores considerados. En la Figura 3 hemos señalado con

la fecha 1 el gradiente de colores que hemos utilizado para construir el gráfico a partir de la serie de Lorenz.

Con la fecha 2 hemos marcado el cuadro correspondiente a los parámetros de la reconstrucción. Por defecto utiliza una dimensión $d_e = 3$ y $T = 1$. Estos valores pueden modificarse (min = 1, máx = 100) moviendo, simplemente, los botones que aparece al lado del cuadrado de dimensión y retardo. Concretamente, para la serie de Lorenz hemos utilizado $d = 2$ y $T = 17$.

El problema a la hora de decidir los parámetros de la reconstrucción es que, en la mayoría de las ocasiones, desconocemos el sistema generador de las series y en consecuencia su dimensionalidad. De los diferentes procedimientos propuestos en la literatura sobre dinámica no lineal para estimar la dimensión de reconstrucción (Abarbanel et al., 1993; Dämmig y Mitschke, 1993) el programa VRA tiene implementado el FNN (False Nearest Neighbors). Este procedimiento consiste en determinar cuándo los puntos en el espacio de fases reconstruido están infinitesimalmente próximos por haber utilizado una dimensión muy pequeña. El aumento progresivo de la dimensión hasta que desaparezcan los falsos infinitésimos proporciona una medida de la dimensión necesaria para la reconstrucción. Para obtener el porcentaje de falsos infinitésimos abrimos el menú "Analysis" y seleccionamos "General Nonlinear Analysis" elegimos "False Nearest Neighbors". El resultado de esta secuencia de elecciones es la ventana que mostramos en la Figura 4. El valor adecuado para la dimensión de reconstrucción es el correspondiente al porcentaje mínimo de falsos infinitésimos. Para la serie de Lorenz, como muestra la figura 4, la dimensión óptima de reconstrucción es 2.

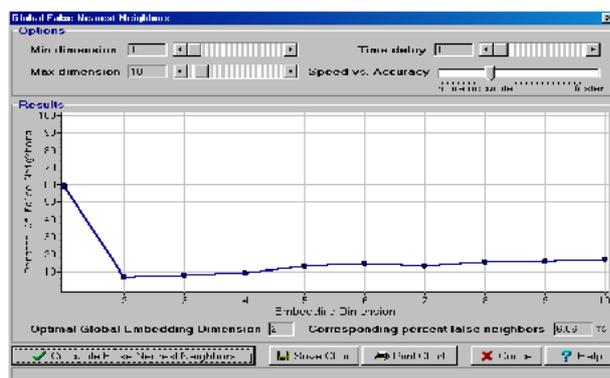


Figura 4. Porcentaje de falsos infinitésimos en función de la dimensión del espacio reconstruido para la serie de Lorenz.

En cuanto al retardo adecuado para reconstruir el atractor contenido en la serie, el programa permite calcular la Información Mutua Media $I(T)$. La información mutua entre observaciones de una serie de tiempo hace referencia a la información que el conocimiento de, por ejemplo, X_i proporciona de X_{i+T} . El valor de $I(T)$, donde T es el retardo, viene dado por

$$I(T) = \sum_{i=1}^N P(x_i, x_{i+T}) \log_2 \left[\frac{P(x_i, x_{i+T})}{p(x_i)P(x_{i+T})} \right] \quad (4)$$

$I(T)$ es una generalización, en dinámica no lineal, de la función de autocorrelación en dinámica lineal. Cuando las medidas son totalmente independientes $P(x_i, x_{i+T}) = P(x_i)P(x_{i+T})$ e $I(T) = 0$ (Abarbanel et al. 1993). En dinámica no lineal se recomienda elegir el valor de T para el que $I(T)$ tenga el primer mínimo. Esta recomendación junto a la de elegir el valor de T correspondiente al primer mínimo de la función de autocorrelación son los dos criterios más utilizados en dinámica no lineal para elegir el retardo adecuado para reconstruir la dinámica del sistema generador de la serie a partir de la misma.

Para obtener la función de información mutua con VRA abrimos nuevamente el menú de "Analysis" y el menú "General Nonlinear Analysis" y ahí seleccionamos "Mutual Infomation". Para la serie de Lorenz que venimos analizando los valores de dimensión de reconstrucción y retardo adecuados serían 2 y 17 respectivamente (ver Figura 5).

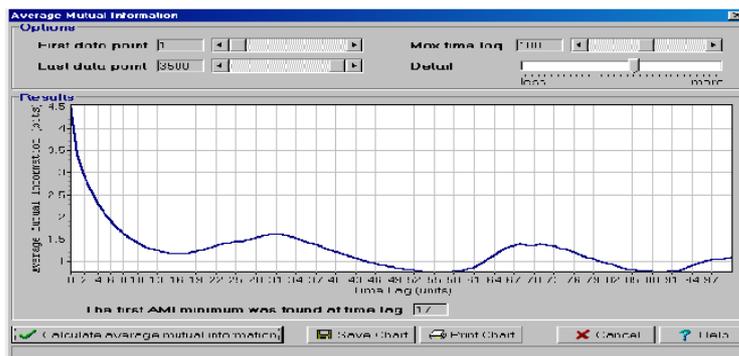


Figura 5. Información mutua media en función del tiempo de retardo para la serie de Lorenz.

Volviendo a la Figura 3, con la flecha 3 hemos marcado los tres métodos para el cálculo de distancias implementados en el programa. Por defecto calcula la distancia euclídea. En este mismo cuadro podemos seleccionar una distancia máxima entre vectores para ser representados por puntos en el gráfico. Moviendo el botón correspondiente al límite superior (Upper bound) seleccionaríamos aquellos vectores cuya distancia esté entre el valor mínimo encontrado en la serie y el valor en el que situemos el límite superior.

Con la flecha 4 hemos marcado el cuadro que permite elegir, de entre 14 archivos diferentes, la combinación de colores que queremos utilizar en el gráfico. Por defecto utiliza el archivo Spectrum.map correspondiente a un gradiente que va de colores cálidos (amarillo) para las distancias pequeñas a colores fríos (azul) para las distancia grandes. En las Figuras 1 y 3 hemos utilizado el archivo Grey.map.

En el gráfico de recurrencia de la Figura 3 se puede apreciar, claramente, una

estructura geométrica incompatible con el comportamiento aleatorio. Tampoco se trata de comportamiento periódico por las diferencias con el gráfico de la Figura 1. Este gráfico corresponde a una serie caótica generada por un sistema de ecuaciones deterministas.

Para obtener los índices correspondientes al RQA abrimos el menú "Analysis" y "Recurrence Plot Analysis" seleccionando "RQA Measurements" obtenemos los valores que aparecen en la Figura 6 para la serie de Lorenz. Es de destacar el alto porcentaje de determinismo incompatible con el comportamiento estocástico.

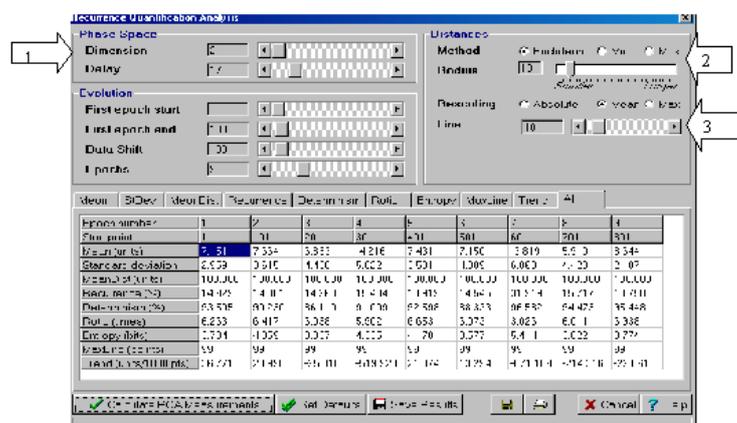


Figura 6. RQA para la serie de Lorenz.

Por defecto VRA divide la serie en trozos (epochs) y para cada uno calcula los índices del RQA. Esta forma de proceder es importante para series largas porque permite identificar transiciones de estado si las hubiera. Se puede, también, calcular los índices de RQA sobre la serie completa.

En el cuadro de diálogo de "Recurrence Quantification Analysis" pueden seleccionarse los parámetros de la reconstrucción (flecha 1) así como la distancia máxima para considerar a dos vectores próximos en porcentajes (flecha 2) y el número de puntos para definir una línea diagonal (flecha 3).

Además de los índices del RQA, el programa VRA tiene implementado un algoritmo que permite calcular la Entropía Espacio-Temporal (STE). STE compara la distribución de distancias del gráfico completo con la distribución de distancias de las líneas diagonales de puntos recurrentes. Proporciona un porcentaje de la máxima entropía. Valores de STE próximos al 100% son indicativos de comportamiento aleatorio, valores cercanos al 0% son índices de comportamiento periódico. Valores entre 0% y 100% indican comportamiento caótico.

En la Figura 7 se muestra el valor de STE para la serie de Lorenz y para una serie generada de una distribución uniforme (0,1). En la Figura 8 se muestra el gráfico de recurrencia de la serie correspondiente a la distribución uniforme.

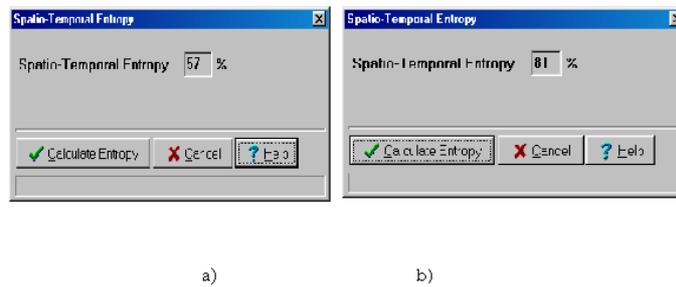


Figura 7. a) STE para la serie de Lorenz. b) STE para la serie uniforme.

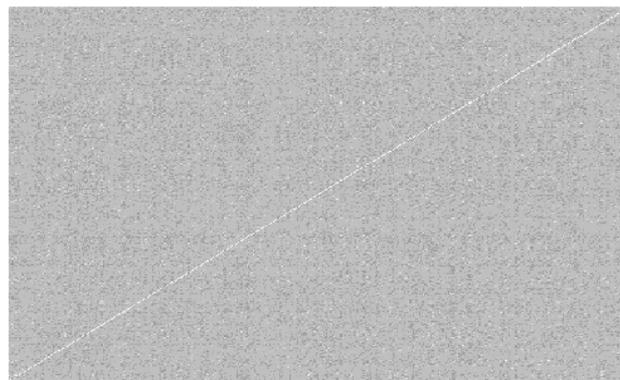


Figura 8. Gráfico de recurrencia para la serie uniforme.

Por cuestiones de espacio no vamos a describir otras utilidades del VRA dado que son similares a las de cualquier programa matemático de propósito general (MATLAB, MATHEMATICA, IMSL) o programa estadístico (SPSS, SAS, BMDP) que tenga implementado un módulo de series temporales.

4.- Referencias.

- Abarbanel, H.D. I., Brown, R., Sidorowich, J.J. y Tsimring, L.S. (1993). The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics*, 65 (4), 1331-1392.
- Barnett, W.A., y Choi, S.S. (1989). A Comparison Between the Conventional Econometric Approach to Structural Inference and the Nonparametric Chaotic Attractor Approach. En W.A. Barnett, J. Geweke y K. Shell (Eds), *Economic Complexity, Chaos, Sunspots, Bubbles and Nonlinearity*. Cambridge: Cambridge University Press.
- Berliner, L.R. (1992). Statistics, Probability and Chaos. *Statistical Science*, 7 (1), 69-90.
- Casdagli, M. (1991). Chaos and Deterministic versus Stochastic Non-linear Modelling. *Journal Royal Statistical Society, B*, 54(2), 303-328.

- Casdagli, M.C. (1997). Recurrence plots revisited. *Physica D*, 108, 12-44.
- Casdagli, M., Eubank, S., Farmer, D.J. y Gibson, J. (1991). State Space Reconstruction in the Presence of Noise. *Physica D*, 51, 52-98.
- Dämmig, M. y Mitschke, F. (1993). Estimation of Lyapunov exponents from time series: the stochastic case. *Physics Letters A*, 178, 385-394.
- Eckmann, J.P., Kamphorst, S.O. y Ruelle, D. (1987). Recurrence Plots of Dynamical Systems. *Europhysics Letters*, 4(9), 973-977.
- Gerr, N.L. y Allen, J.C. (1993). Stochastic Versions of Chaotic Time Series: Generalized Logistic and Hénon Time Series Models. *Physica D*, 68, 232-249.
- López, A.M. (1998). *Aproximación cualitativa al estudio de la dinámica no lineal. Aplicaciones al análisis de series temporales: un estudio de simulación*. Tesis Doctoral. Universidad de Sevilla.
- López, A.M., Camacho, C. y García, A.R. (2001). Caracterización cualitativa de series de tiempo: geometría y dinámica en el espacio de fases reconstruido. *Metodología de las Ciencias del Comportamiento*. (En prensa).
- Mañé, R. (1981). Dynamical systems and turbulence. En D.Rand y L.S. Young (Eds), *Lecture Notes in Mathematics*. No. 898. Berlin: Springer.
- McCaffrey, D., Ellner, S., Gallant, A.R. y Nychka, D. (1992). Estimating the Lyapunov Exponent of a Chaotic Systems with Nonparametric Regression. *Journal of the American Statistical Association*, 87(419), 682-694.
- Montero, F. y Morán, F. (1992). *Biofísica. Procesos de Autoorganización en Biología*. Madrid: EUDEMA Universidad.
- Nychka, D., Ellner, S., Gallant, A.R. y McCaffrey, D. (1992). Finding Chaos in Noisy Systems. *Journal Royal Statistical Society*, 54(2), 399-426.
- Packard, N.H., Crutchfield, J.P., Farmer, J.D. y Shaw, R.S. (1980). Geometry from a Time Series. *Physical Review Letters*, 45(9), 712-717.
- Sayers, C.L. (1991). Statistical Inference Based Upon Non-Linear Science. *European Economic Review*, 35, 306-312.
- Sugihara, G. y May, R.M. (1990). Nonlinear Forecasting as a Way of Distinguishing Chaos from Measurement Error in Time Series. *Nature* 344, 734-741.
- Takens, F. (1993). Detecting nonlinearities in stationary time series. *International Journal of Bifurcation and Chaos*, 3(2), 241-256.

Trulla, L.L, A. Giuliani, J.P. Zbilut, and C.L. Webber, Jr. (1996). Recurrence quantification analysis of the logistic equation with transients. *Physics Letters A*, 223, 255-260.

Zbilut, J.P. y Webber, C.L. Jr. (1992). Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, 171, 199-203