

Gráficos de Recurrencia Simbólicos

M. VICTORIA CABALLERO; MARIANO MATILLA Y MANUEL RUIZ

Universidad de Murcia; UNED; UPCT;

mvictori@um.es;mmatilla@cee.uned.es;manuelruiz.spain@gmail.com

Resumen

Un gráfico de recurrencia simbólico de una serie vectorial de \mathbb{R}^m es una representación en el plano de los símbolos recurrentes de su correspondiente serie de símbolos de S_m . Este gráfico puede ser coloreado si en su representación distinguimos con un color distinto los diferentes símbolos recurrentes de S_m .

En este trabajo, además de visualizar gráficos de recurrencia simbólicos coloreados, se definen medidas sobre estos gráficos que cuantifican las distintas estructuras que están presentes y que toman valores en rangos distintos según la dinámica de la serie que se estudie. Una de estas medidas es *la proporción de puntos recurrentes* que toma valores significativamente distintos si se calcula para el gráfico de recurrencia simbólico de una serie dependiente o para el de una serie independiente.

1. Introduction

Los gráficos de recurrencia y sus análisis cuantitativo son una herramienta útil para estudiar y comprender el comportamiento de un sistema dinámico a través de series temporales generadas por el mismo, y son especialmente interesantes cuando el sistema generador de la serie es no lineal.

Los gráficos de recurrencia fueron propuestos por Eckmann et al. ([2]) y para construirlos, es preciso, dada una serie temporal real o vectorial, fijar el valor de un parámetro $\varepsilon > 0$; de modo que estos gráficos pueden variar según este valor fijado a priori. Para cuantificar las estructuras más importantes que aparecen en un gráfico de recurrencia, Zbilut y Webber [5, 4] y más tarde Marwan [3] desarrollaron una serie de medidas recogidas en el *Análisis Cuantitativo de la Recurrencia* (RQA). Estas medidas han sido usadas por los investigadores para detectar distintos comportamientos dinámicos a partir de las estructuras recurrentes existentes en estos gráficos.

En este trabajo, nosotros trabajamos con los gráficos de recurrencia simbólicos, nos interesa cuándo y cuántas veces las componentes de los elementos de una serie temporal vectorial están ordenadas según el mismo patrón y qué estructuras y cuántas aparecen en estos gráficos. Para ello, comenzamos fijando la notación básica que vamos a utilizar para pasar en la sección tercera a definir matriz simbólica coloreada y su visualización en el plano, es decir, su gráfico de recurrencia simbólico (coloreado). Además, ilustramos con distintos ejemplos cómo estos gráficos de recurrencia simbólicos, en particular los coloreados, nos informan del comportamiento dinámico de la serie temporal vectorial que representan. Se definen medidas a partir de la información proporcionada por

estos gráficos en la sección cuarta para pasar a utilizarlas con el fin de distinguir cambios en el modelo generador de los datos que se estudian.

2. Preliminares y notación

Para un entero positivo $m \geq 2$ denotamos por S_m el grupo simétrico de orden $m!$, esto es el formado por todas las permutaciones del conjunto $\{1, 2, \dots, m\}$. Cada elemento $\pi = (i_1, i_2, \dots, i_m) \in S_m$ es un símbolo.

Sea un subconjunto $X \subseteq \mathbb{R}^m$, se define la aplicación:

$$s : X \rightarrow S_m$$

que asigna a cada elemento $\bar{x} \in X$ el símbolo de S_m que da el orden de sus componentes en el conjunto de los números reales, y tal que si dos o más componentes de \bar{x} son iguales el orden viene dado por su orden de colocación en la m -úpla. El símbolo asignado a \bar{x} se dice que es su tipo.

Definición: Se dice que un vector de $\bar{x} \in \mathbb{R}^m$ es de tipo $\pi \in S_m$, si y solo si, $\pi = (i_1, i_2, \dots, i_m)$ es el único símbolo que satisface las siguientes condiciones:

- a) $x_{i_1} \leq x_{i_2} \leq \dots \leq x_{i_m}$ y
- b) $i_{s-1} < i_s$ if $x_{i_{s-1}} = x_{i_s}$.

La condición (b) garantiza la unicidad del símbolo π asignado al vector \bar{x} . La aplicación s transforma cada $\bar{x} \in \mathbb{R}^m$ en un símbolo.

Se define la función indicador $I : X \times X \rightarrow S_m$ como

$$I(\bar{x}, \bar{y}) = \begin{cases} 1 & \text{if } s(\bar{x}) = s(\bar{y}) \\ 0 & \text{en otro caso} \end{cases}. \quad (1)$$

esta función $I(\bar{x}, \bar{y})$ vale 1 siempre que el orden de las componentes de \bar{x} e \bar{y} sea el mismo.

En particular, si $\{x_t\}_{t=1}^T$ es una serie temporal unidimensional, que es la realización de un proceso estocástico real, y $m \geq 2$ se puede construir una serie vectorial cuyos elementos son de la forma

$$\bar{x}_t = (x_t, x_{t+1}, \dots, x_{t+(m-1)}), \quad (2)$$

se dice que la serie temporal $\{\bar{x}_t\}_{t=1}^n$ se ha obtenido al embeber la serie temporal unidimensional en un espacio m -dimensional (\mathbb{R}^m), con $n = T - m + 1$, donde cada \bar{x}_t es una m -historia y m es la dimensión embedding. Luego si $X = \{\bar{x}_t\}_{t=1}^n$ es la serie vectorial embebida, aplicación s transforma la serie vectorial de m -historias en una sucesión de símbolos.

3. Representando la función indicador $I(\cdot, \cdot)$: Gráfico de recurrencia simbólico

Dada una serie temporal vectorial en \mathbb{R}^m , la función indicador $I(\cdot, \cdot)$ informa sobre cuándo y cuántas veces la serie temporal visita cada una de las $m!$ zonas de \mathbb{R}^m determinadas por los $\binom{m}{2}$ hiperplanos, definidos por la igualdad de las componentes tomadas de dos en dos ($\{(x_1, \dots, x_n) \in \mathbb{R}^m : x_i = x_j\}$).

Si tenemos una serie temporal vectorial $\{\bar{x}_i\}_{i=1}^n$, con $\bar{x}_i \in \mathbb{R}^m$, al igual que para los gráficos de recurrencia clásicos, se puede definir una matriz $S \in M_n$ donde

$$S_{ij} = I(\bar{x}_i, \bar{x}_j) = \begin{cases} 1 & \text{si } s(\bar{x}_i) = s(\bar{x}_j) \\ 0 & \text{en otro caso} \end{cases}$$

con $i, j = 1, 2, \dots, n$. Cuando $S_{ij} = 1$, se dice que \bar{x}_i y \bar{x}_j son *recurrentes*. Esta es la matriz de recurrencia simbólica correspondiente a la serie vectorial $\{\bar{x}_i\}_{i=1}^n$, que puede ser representada en el plano, donde se coloca un punto en la posición (i, j) cuando $S_{ij} = 1$ y cuando $S_{ij} = 0$, la posición (i, j) estará vacía. Un gráfico de recurrencia simbólico supone la representación de la función indicador $I(\cdot, \cdot)$.

Si distinguimos entre los distintos símbolos estos gráficos pueden colorearse asignando a cada símbolo un color distinto. En este caso la matriz simbólica *coloreada* se definiría como:

$$S_{ij} = I(\bar{x}_i, \bar{x}_j) = \begin{cases} s_i & \text{si } s(\bar{x}_i) = s(\bar{x}_j) \\ 0 & \text{en otro caso} \end{cases}$$

donde s_i es el número asignado al símbolo $s(\bar{x}_i)$, con $i, j = 1, 2, \dots, n$ y cuya representación en el plano situaría un punto en (i, j) , cuyo color dependerá del símbolo $S_{ij} = s_i$ y que cuando $S_{ij} = 0$, la posición (i, j) estará vacía. Estos gráficos se llaman *gráficos de recurrencia simbólicos coloreados*.

Ejemplos de la utilidad de los gráficos de recurrencia simbólicos se muestran a continuación. Para ello, a partir de una serie temporal real, elegido un valor de la dimensión embedding m , se ha embebido la serie real en \mathbb{R}^m construyendo una serie vectorial. Este procedimiento ha sido aplicado a varias series reales generadas por distintos modelos dinámicos. En todos los casos se ha partido de una serie temporal real de 1000 datos y se ha construido la serie vectorial en \mathbb{R}^4

Para obtener la Figura 1 se ha utilizado una serie temporal procedente de un sistema dinámico tridimensional continuo y autónomo ([1]), se trata de la componente x de una órbita periódica. A partir de ella, se obtiene la vectorial y su correspondiente serie temporal de símbolos. Podemos observar la importancia del color, en la Figura 1(a) se utiliza el mismo color para todos los símbolos recurrentes, mientras que en la Figura 1(b) se usan diferentes colores para los diferentes símbolos recurrentes y el gráfico indica que la serie de símbolos está mayoritariamente formada por dos símbolos distintos.

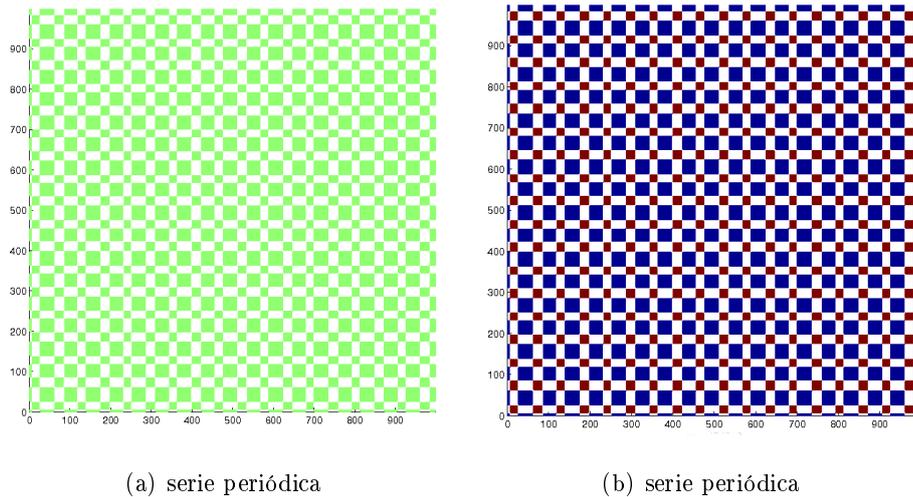


Figura 1: Gráficos de recurrencia simbólico correspondientes a una serie periódica. En la Figura 1(a) a todos los símbolos recurrentes se le asocia el mismo color. En la Figura 1(b) distintos símbolos recurrentes, distinto color.

La Figura 2(a) muestra el gráfico de recurrencia simbólico para una serie temporal i.i.d ($\mathcal{N}(0, 1)$) y la Figura 2(b) es el gráfico de recurrencia simbólico de una serie temporal AR(1).

Análogamente, La Figura 3(a) muestra el gráfico de recurrencia simbólico para una órbita del sistema dinámico ($[0, 1], f(x) = 4x(1 - x)$). Por último, para el gráfico de la Figura 3(b), hemos generado una serie a partir del sistema dinámico ($[0, 1], f(x) = ax(1 - x)$) variando el parámetro a cada cien datos.

A diferencia de los gráficos de recurrencia clásicos ([2]), en los que a priori hay que fijar un valor $\varepsilon > 0$ y una dimensión embedding m para embeber la serie real, de modo que dos m -historias son recurrentes cuando están a una distancia menor que $\varepsilon > 0$, en un gráfico de recurrencia simbólico solo hay que fijar la dimensión embedding a priori puesto que la recurrencia de dos m -historias ocurre cuando tienen el mismo patrón de orden en sus componentes.

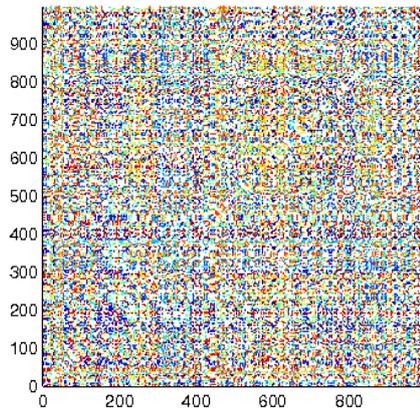
4. Medidas en gráficos de recurrencia simbólicos

En esta sección se van a definir distintas medidas relacionadas con los gráficos de recurrencia simbólicos de una serie temporal vectorial.

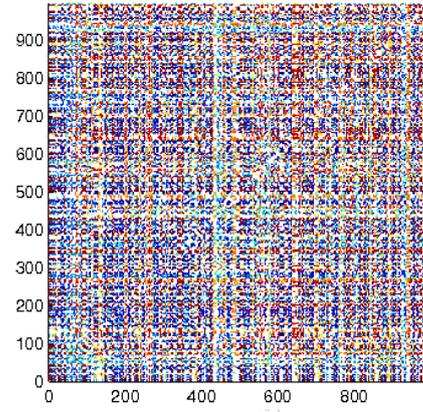
- **Proporción de puntos recurrentes en un gráfico de recurrencia simbólico:**

Se trata de la proporción de parejas de elementos de la serie vectorial que son recurrentes:

$$RECS = \frac{1}{n^2} \sum_{i,j=1}^n S_{ij},$$

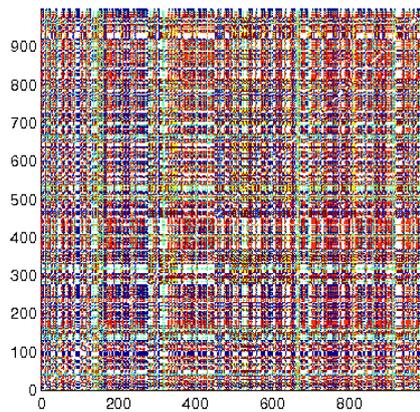


(a) serie $N(0,1)$

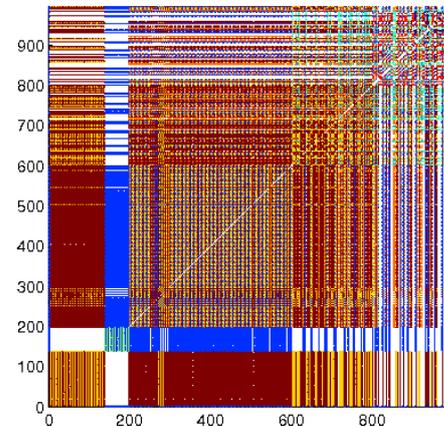


(b) serie AR(1)

Figura 2: La Figura 2(a) es el gráfico de recurrencia simbólico correspondiente a una serie iid y la Figura 2(b) es el gráfico correspondiente a un modelo autorregresivo $x_t = 0,8x_{t-1} + \epsilon_t$.



(a) logística, parámetro=4



(b) logística, parámetro variable

Figura 3: Gráfico de recurrencia simbólico para $x_t = ax_{t-1}(1 - x_{t-1})$ con condición inicial $x_0 = 0,15$

Es una medida de la densidad de los pares de m -historias que tienen el mismo símbolo (son del mismo tipo) de entre la totalidad de los pares de elementos de la serie vectorial. Esta definición se corresponde con la del estimador de la correlación integral simbólica.

La Tabla 1 muestra el valor de esta proporción de puntos recurrentes en los gráficos de recurrencia simbólicos de la sección anterior y también en los gráficos cuando se realiza el embedding de las series reales en \mathbb{R}^3 y en \mathbb{R}^4 :

Tabla 1: Proporción de puntos recurrentes en un gráfico de recurrencia simbólico

Modelo	$RECS, m = 3$	$RECS, m = 4$
Periódica (Figura 1)	0,4915	0,4576
$N(0, 1)$ (Figura 2(a))	0,1667	0,0417
$AR(1)$ (Figura 2(b))	0,1774	0,0523
$x_{t+1} = 4x_t(1 - x_t)$ (Figura 3(a))	0,2566	0,1162

Los valores obtenidos para las series deterministas, correspondientes a las Figuras 1 y 3(a), son sensiblemente mayores que los correspondientes a las series no deterministas. En relación a estas últimas, también podemos observar que mientras los valores de la serie i.i.d. están cercanos a $1/3!$ y $1/4!$ para $m = 3$ y $m = 4$, respectivamente, los correspondientes a la serie $AR(1)$ son ligeramente mayores.

▪ **Proporción de puntos recurrentes asignados a un símbolo:**

Se trata de medir la densidad de la recurrencia de un símbolo de S_m , respecto del total de pares de la serie vectorial. Si llamamos π a un símbolo cualquiera de S_m

$$RECS(\pi) = \frac{1}{n^2} \sum_{i,j=1}^n S_{ij}(\pi),$$

donde solo se tienen en cuenta los $S_{ij} = 1$ que tienen asignado 1 cuando $s(\bar{x}_i) = s(\bar{x}_j) = \pi$.

Basta para calcular esta medida utilizar los gráficos de recurrencia de las series utilizadas en la sección anterior embebidas en \mathbb{R}^3 , debido a que en S_3 hay 6 símbolos mientras que, si utilizamos los gráficos de recurrencia de la serie embebida en \mathbb{R}^4 el número de símbolos es 24.

Los resultados que se recogen en la Tabla 2 nos indican que si la serie es periódica, tal como nos mostraba la Figura 1(b), la mayor proporción de recurrencias ocurren en 3-historias cuyos tipos son los correspondientes a los símbolos denotados como π_1 y π_6 , y el número de estados recurrentes de los restantes tipos son muy pocos. Si nos fijamos en los resultados correspondientes a la Figura 3(a), vemos que no aparece uno de los símbolos, siendo las 3-historias del tipo π_1 la que presentan, con diferencia, el mayor número de estados recurrentes.

Tabla 2: Tasa de recurrencia simbólica de cada uno de los símbolos

	Periódica (Figura 1)	$N(0, 1)$ (Figura 2(a))	$AR(1)$ (Figura 2(b))	$x_{t+1} = 4x_t(1 - x_t)$ (Figura 3(a))
$RECS(\pi_1)$	0,3564	0,031	0,0484	0,1372
$RECS(\pi_2)$	$9,0452 \cdot 10^{-10}$	0,0256	0,0207	0,0036
$RECS(\pi_3)$	$3,0151 \cdot 10^{-5}$	0,0275	0,0179	0,0156
$RECS(\pi_4)$	$1,3266 \cdot 10^{-4}$	0,0243	0,0190	0,0357
$RECS(\pi_5)$	$5,628 \cdot 10^{-5}$	0,0265	0,0161	0,0646
$RECS(\pi_6)$	0,1343	0,0313	0,0553	0

Para las Figuras 2(a) y 2(b) los resultados muestran que mientras que cuando la serie es i.i.d. la recurrencia a los distintos símbolos es más o menos igual, en el caso del la serie generada por un modelo AR(1), alguno de los símbolos llega a cuadruplicar las recurrencias del símbolo que menos recurrencias tiene.

■ **Líneas verticales:**

Se trata de la proporción de puntos recurrentes que forman líneas verticales en un gráfico de recurrencia simbólico.

Una línea vertical de longitud v en un gráfico de recurrencia simbólico, que comienza en el momento t correspondiente al estado \bar{x}_t cuyo tipo viene dado por el símbolo $s(\bar{x}_t)$, ocurre cuando $S_{ij+k} = 1$ para $k = 0, 1, \dots, v - 1$; además la línea tiene el mismo color (análogamente se podrían definir líneas horizontales). Se tiene una línea vertical cuando al menos hay dos símbolos recurrentes formándola, luego la longitud mínima para considerar línea vertical sera $v_{\text{mín}} = 2$.

La Tabla 3 muestra los valores de esta medida para los Figuras de la Sección 3. Se observa que los valores correspondientes a las series dependientes son sensiblemente mayores que el correspondiente a la serie i.i.d. Además, estas estructuras son mayores cuanto más determinista es la serie (entendiendo por más determinismo cuando el horizonte de predicción es mayor), por ello es mayor cuando la serie es periódica que cuando es una órbita caótica.

$$VS = \frac{\sum_{v=v_{\text{mín}}}^n vP(v)}{\sum_{i,j=1}^n S_{ij}},$$

■ **Longitud media de las líneas verticales:**

$$TS = \frac{\sum_{v=v_{\text{mín}}}^n vP(v)}{\sum_{v=1}^n P(v)},$$

Tabla 3: Proporción de puntos recurrentes que forman líneas verticales

Modelo	VS
Periódica (Figura 1)	0,9985
$N(0, 1)$ (Figura 2(a))	0,03807
$AR(1)$ (Figura 2(b))	0,2283
$x_{t+1} = 4x_t(1 - x_t)$ (Figura 3(a))	0,284

y da el tiempo seguido medio que la serie vectorial permanece en una de las $m!$ porciones en las que queda dividido \mathbb{R}^m cuando se considera la serie de símbolos. En la Tabla 4 podemos

Tabla 4: Longitud media de las líneas verticales

Modelo	TS
Periódica (Figura 1)	26,097
$N(0, 1)$ (Figura 2(a))	2,331
$AR(1)$ (Figura 2(b))	2,761
$x_{t+1} = 4x_t(1 - x_t)$ (Figura 3(a))	3,125

ver los valores que toma esta medida en los gráficos de las series de la sección 3. Los resultados indican que la serie vectorial correspondiente a la serie real periódica tiene el mayor tiempo medio de permanencia en una de las zonas de la partición hecha de \mathbb{R}^4 por el conjunto de símbolos S_4 y el menor le corresponde a una serie i.i.d.

5. Gráfico de recurrencia simbólico y cambio estructural

Teniendo en cuenta cómo se comportan las medidas vistas en la sección anterior sobre distintos tipos de datos, éstas podrían ayudarnos a distinguir cambios estructurales en la dinámica del sistema del cual procede la serie que se estudie.

Para ello hemos generado dos series de longitud 1000 obtenidas de dos modelos cuyos parámetros han cambiado en algún momento a lo largo del periodo temporal y hemos procedido de la siguiente manera: A partir de la serie unidimensional se han tomado subseries de longitud fijada k , tal que dos subseries consecutivas tienen en común un número de elementos también fijado, cada subserie se ha embebido en \mathbb{R}^4 y hemos calculado la proporción de puntos recurrentes del gráfico de recurrencia simbólico correspondiente a cada una de las subseries, de modo que cuando haya un cambio en el valor de ésta será debido a un cambio en los parámetros del modelo de que proceden los datos.

El primer conjunto de datos que se ha utilizado ha sido el correspondiente a la Figura 3(b). La serie de 1000 datos se ha generado a partir del sistema dinámico determinista $([0, 1], f(x) = ax(1-x))$, variando el parámetro a , que toma los valores $3 + 0,1 \cdot i$ con $i = 1, 2, \dots, 10$. De esta serie se van tomando subseries de 100 elementos, de modo que dos subseries sucesivas tienen en común 90 elementos, y se construye la serie vectorial embebida en \mathbb{R}^4 . Para cada uno de estos gráficos de recurrencia se calcula la proporción de puntos recurrentes. La Figura ?? informa que los parámetros del sistema generador de datos han debido de cambiar puesto que la dinámica del mismo cambia puesto que la proporción de puntos recurrentes cambia sensiblemente.

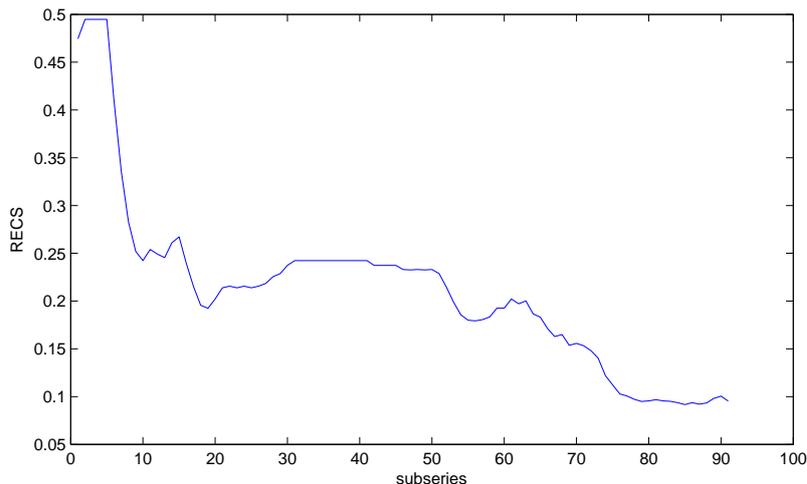


Figura 4: Evolución de la medida RECS

Hemos procedido de modo análogo con una serie real de longitud 1000, de los que los primeros 500 datos han sido generados con el modelo $x_t = 0,8x_{t-1} + e_t$ con una condición inicial aleatoria y donde $e_t \sim N(0, 1)$. Los otros 500 datos de la serie se generan con el modelo $x_t = 0,3x_{t-1} + e_t$. En la Figura se puede observar el comportamiento de la proporción de puntos recurrentes de los gráficos de recurrencia en \mathbb{R}^4 de subseries de longitud 100. Las primeras subseries vectoriales están formadas por elementos de la serie real generados por el primer modelo, vemos que a partir del valor 41-42 de RECS hay un cambio significativo en los valores de esta medida (ya los últimos elementos de la subserie vectorial 42 están generados por el segundo modelo). Todo ello puede observarse en la Figura 5

6. Conclusiones

Este es un trabajo descriptivo para estudiar series temporales y se encuentra en la fase inicial. Sin embargo, por una parte tenemos que resaltar que los gráficos de recurrencia simbólicos informan

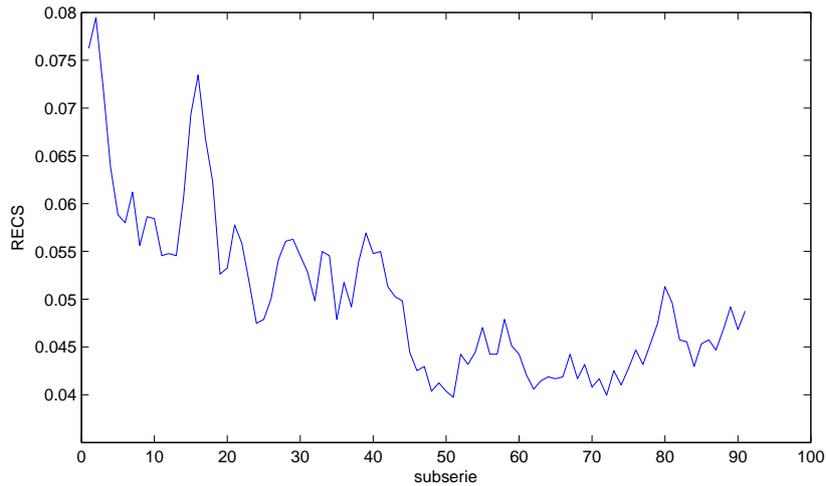


Figura 5: Evolución de la medida RECS

sobre la dinámica de la serie real y no es preciso fijar un valor a priori ε necesario en los gráficos de recurrencia clásicos y por otra, los resultados obtenidos en las secciones 4 y 5 nos invitan a seguir profundizando en estos gráficos de recurrencia y en las medidas elaboradas. También tenemos que señalar que en la última sección solo hemos utilizado la medida proporción de puntos recurrentes para detectar cambios de dinámica y es posible que otras medidas puedan dar mejores resultados. Para finalizar diremos que tanto los gráficos de recurrencia simbólicos como estas medidas serán aplicadas para estudiar otros aspectos de las serie temporales, como causalidad y sincronización de series.

Referencias

- [1] Balibrea, F., Caballero, M.V. and Molera, L. (2008) Recurrence quantification analysis in Liu's attractor. *Chaos, Solitons & Fractals*, 36, 3, 664-670.
- [2] Eckmann J, Kamphorst S. and Ruelle D. (1987) Recurrence plots of dynamical systems. *Europhysics Letters*, 5, 973-977.
- [3] Marwan N. Encounters with neighbours-Current Developments of Concepts Based on Recurrence Plots and their Applications-. PhD Thesis, Institute of Physics, Postdam: University of Postdam; 2003.
- [4] Webber CL and Zbilut JP. Dynamical assessment of physiological systems and states using recurrence plot strategies. *J Appl Physiol* 1994;76(2):965-973.

- [5] Zbilut JP and Webber CL. Embeddings and delays as derived from quantification of recurrence plots. *Phys Lett A* 1992;171:199-203.