

UNIVERSIDAD NACIONAL DE COLOMBIA

Aplicación de la teoría de caos para el análisis y pronóstico de series de tiempo financieras en Colombia

Maestría en Ingeniería Administrativa

Autor: Ing. Juan Pablo Gallego Valencia

25/08/2010

Director: Santiago Medina Hurtado

Se introducen algunas técnicas de análisis de series de tiempo financieras basadas en el paradigma de caos determinista tales como el exponente de Hurst, la teoría de embebimiento y el cálculo de invariantes no lineales. Se desarrolla un análisis con el fin de detectar dinámicas no lineales en la serie de la tasa representativa del mercado colombiano (TRM) y se ajusta un modelo para su pronóstico.

Tabla de contenido

Tabla de contenido	3
1. Motivación.....	5
2. Conceptos básicos.....	7
2.1 Sistema dinámico determinista.....	7
2.2 Espacio de fase de un sistema dinámico.....	7
2.3 Atractor extraño.....	7
2.4 Comportamiento caótico.....	8
2.5 Cantidades Invariantes.....	9
3. Herramientas de análisis.....	11
3.1 Análisis de re-escalamiento y el exponente de Hurst.....	11
3.1.1 El método.....	11
3.1.2 Interpretación.....	13
3.2 Teoría de embebimiento.....	16
3.2.1 Teorema de Takens.....	17
3.2.2 Selección del tiempo de rezago.....	18
3.2.3 Selección de la dimensión de embebimiento.....	22
3.3 Método de de pronóstico.....	26
4. Invariantes Topológicos en Sistemas No lineales.....	31
4.1 Entropía de la información.....	31
4.2 Exponentes de Lyapunov.....	33
4.3 Dimensión de Correlación.....	36
5. Análisis estadístico por medio del método <i>Surrogate Data</i>	41
5.1 Pruebas estadísticas.....	42
5.1.1 Prueba Kolmogorov – Smirnov.....	43
5.1.2 Prueba χ^2	44
6. Análisis de datos.....	45
6.1 Análisis de la tasa representativa del mercado (TRM).....	45
6.1.1 Los datos.....	45
6.1.2 Análisis de reescalamiento (R/S).....	48

6.1.3	Teoría de embebimiento.....	49
6.1.4	Cálculo de invariantes.....	53
6.1.5	Análisis Lineal por medio de Surrogate Data.....	60
	Conclusiones y discusión.....	62
	Bibliografía	64

1. Motivación

Desde la utilización del movimiento Browniano Geométrico, descrito por Einstein en 1910, para modelar el comportamiento de los precios de las tasas de cambio (Peters 1994), cuando inicialmente fue desarrollado para describir el movimiento errático de una partícula suspendida en un fluido; hasta la utilización de emulaciones básicas del funcionamiento cerebral, como los modelos de redes neuronales artificiales, para replicar las dinámicas que rigen el comportamiento de precios de acciones, la necesidad de comprender los fenómenos que gobiernan las fluctuaciones del mercado de capitales en cualquier economía ha llevado a aplicar herramientas de análisis derivadas de campos como la física estadística, la dinámica de fluidos, la teoría de la información, el procesamiento de señales, entre otras, que a simple vista no tienen ningún tipo de relación con las finanzas (Kantz y Schreiber 2006).

Uno de los campos de estudio cuyas aplicaciones se han explorado en el área financiera es la teoría de sistemas dinámicos no lineales, y aunque este campo de estudio cuenta con un soporte matemático enorme (Eckmann y Ruelle 1985), el estudio de series de tiempo que se consideran derivadas de este tipo de comportamiento se considera aún en una fase experimental (Sprott 2003). Sin embargo se puede observar un gran interés por mejorar los métodos de análisis que se han desarrollado a lo largo los últimos treinta años (Grassberger y Procaccia 1983) (Broomhead y King 1985) (Wolf, et al. 1985) (Judd 1992) (Abarbanel 1996) (Díaz 2000) (Small 2005) (Kantz y Schreiber 2006) (Small 2005).

Desde el descubrimiento de la dinámica irregular (que ahora llamamos Caos Determinista) por parte de Poincaré a principios de 1900, la cantidad de contribuciones al campo de los sistemas dinámicos no lineales ha crecido casi de manera exponencial, y con mayor razón ahora que se cuenta con una capacidad de procesamiento computacional enorme en comparación con la del siglo anterior. Podría decirse que es posible “naufregar” en la búsqueda de información relacionada con las herramientas que se presentarán en este trabajo debido a que sus aplicaciones abarcan tantas disciplinas científicas que es casi imposible cubrir todas las contribuciones realizadas a las metodologías.

Los precios de los instrumentos financieros cambian de manera errática y no siempre existirá certeza sobre el origen de este comportamiento, es decir, si estas variables en realidad pueden catalogarse como aleatorias o presentan algún tipo de dinámica determinista embebida. De presentarse este último caso, se busca ajustar esta estructura dinámica por medio de un modelo, el cual serviría en la descripción del fenómeno y en su pronóstico (Nakamura y Small 2006).

Se han encontrado gran variedad de trabajos internacionales que buscan modelar dinámicas no lineales e incluso comportamiento caótico en series de tiempo financieras, algunos ejemplos se observan en (Cecen y Erkal 1996), (Matsuba, y otros 2000), (Strozzi y Zaldívar 2005). Los resultados son variables y de hecho discutibles. Sin embargo casi todos estos estudios coinciden en la presencia de estructuras dinámicas que no pueden ser modelables por medio de las herramientas de análisis lineal que, de ser posible su modelamiento, pueden contribuir al mejoramiento de los pronósticos.

En el caso colombiano, en el estudio de las series financieras prevalece el análisis desde el punto de vista econométrico y estocástico (Giraldo 2005) (Vélez y Melo 1999) (Misas, y otros 2005) (Ospina y Cerezo 2008), por medio de herramientas heurísticas (Mejía 2008) y de inteligencia artificial (Velásquez y Montoya 2005) (Jalil y Misas 2007) (Santana, y otros 2006). En la revisión realizada solo se encontró un trabajo escrito por autores argentinos y brasileros, el cual utiliza algunas herramientas de análisis no lineal de series de tiempo, en busca de multi-fractalidad en las series de los índices del mercado Latinoamericano, incluyendo el índice general de la bolsa de valores de Colombia (IGBC) (Zunino, y otros 2009). Este estudio concluye positivamente sobre la existencia de las estructuras multi-fractales en la serie del IGBC, lo cual fortalece la motivación para el desarrollo de esta investigación.

Luego el interés fundamental de este trabajo, es realizar un acercamiento inicial a las metodologías utilizadas para el análisis de series de tiempo, de manera que pueda determinarse la presencia o ausencia de dinámicas no lineales embebidas en su comportamiento, bajo el paradigma del caos determinista. A partir de este análisis se buscará aplicar estas metodologías en un caso particular del mercado colombiano, como es el precio de la tasa representativa del mercado TRM.

El documento está organizado de la siguiente manera: en la sección 2 se introducirán algunos conceptos básicos necesarios para la comprensión del trabajo desarrollado. La sección 3 y 4 profundizarán sobre las herramientas desarrolladas para el análisis y modelado del comportamiento no lineal, lo cual se complementará con algunos análisis estadísticos basados en la metodología de *Surrogate Data* mostradas en la sesión 5. Finalmente, se presenta la aplicación particular del análisis de la tasa representativa del mercado (TRM) junto con algunas conclusiones y discusiones generadas a partir de los resultados obtenidos, en las secciones 6 y 7 respectivamente.

2. Conceptos básicos

En este capítulo se describirán algunos conceptos básicos que serán necesarios para la correcta interpretación de la investigación consignada en esta monografía. Cabe anotar que a medida que se desarrolle el trabajo se introducirán nuevos conceptos y de ser necesario se profundizará en los que se presentan en esta sección. Por lo pronto se abarcarán algunas definiciones sobre sistemas dinámicos y el paradigma del caos determinista.

2.1 Sistema dinámico determinista

Por sistema dinámico se entiende un conjunto de elementos interconectados entre sí y destinados a cumplir una tarea específica. Este conjunto de elementos, o sistema, describe un comportamiento cuya evolución depende del tiempo, ya sea de forma explícita o implícita y de forma continua o discreta. Se dice que es determinista porque los valores futuros del sistema pueden ser determinados a partir de los valores actuales. En el caso en que el sistema dependa de una variación temporal discreta también suele denominarse Mapa discreto.

Los sistemas dinámicos se pueden clasificar en dos tipos: lineales, cuando el sistema cumple con el principio de superposición (las salidas del sistema pueden describirse como una combinación lineal de sus entradas); y no lineales, cuando el sistema no cumple con el principio de superposición. El número de variables de las cuales depende el sistema es llamado Grados de Libertad o Estados del Sistema.

2.2 Espacio de fase de un sistema dinámico

Se le llama espacio de fase o espacio de estados (de un sistema completamente determinista) al espacio vectorial en el cual un punto representa un estado del sistema y viceversa. En este espacio los ejes coordenados son los estados del sistema dinámico. Por otro lado el número de estados es igual a la dimensión del sistema y también se denomina grados de libertad del sistema (Figura 1).

2.3 Atractor extraño

De forma general se puede decir que un atractor es una región o conjunto cerrado en el espacio de fase al cual convergen las trayectorias de un sistema, entendiéndose por trayectoria la evolución temporal del sistema a partir de una condición inicial específica. Los atractores cumplen las siguientes propiedades (Strogatz 2000):

- El atractor es invariante, cualquier trayectoria que comience dentro del atractor, permanecerá en él en todo momento (siempre y cuando no se perturbe el sistema).

- Al atractor convergen las trayectorias desde un conjunto abierto de condiciones iniciales, al cual se le llama la “cuenca de atracción”.
- El atractor es el mínimo conjunto que cumpla las dos condiciones anteriores, es decir, no contiene subconjuntos propios con sus mismas propiedades

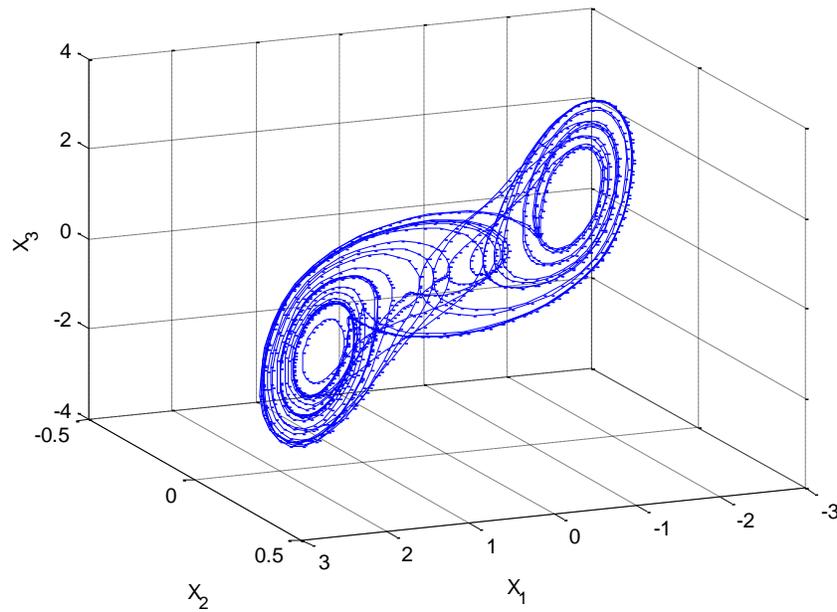


Figura 1: Espacio de fase conformado por los tres estados del sistema dinámico de Chua en el cual se muestra su atractor extraño cuando el sistema está en régimen caótico.

Por atractor extraño se entiende una región del espacio de fase donde la evolución del sistema dinámico está confinada en una región acotada y sin embargo las trayectorias cercanas del atractor se alejan exponencialmente, teniendo en cuenta que la distancia máxima que pueden separarse está limitada por el tamaño del atractor (Figura 1).

2.4 Comportamiento caótico

En sistemas dinámicos, se conoce como caos a un comportamiento aperiódico de un sistema determinístico. Se dice que el comportamiento es aperiódico dado que el sistema no tiende a un punto fijo ó a una órbita periódica en el plano de fase cuando $t \rightarrow \infty$. Por largo tiempo se consideró que el comportamiento caótico era una forma de ruido, debido a que las herramientas con las cuales pretendía estudiarse eran esencialmente lineales.

La principal herramienta para el análisis lineal de señales es la transformada de Fourier, y fue diseñada para hallar la descomposición en senos y cosenos de una serie de datos observados. Al implementar un análisis de Fourier sobre una serie de tiempo de origen caótico, el espectro de frecuencias mostrará un resultado muy semejante al de un ruido. Esto se debe a que el caos se caracteriza por presentar un ancho de banda amplio en el dominio de la frecuencia con algunos picos en franjas estrechas (Abarbanel 1996).

Este comportamiento aperiódico presenta una gran sensibilidad a cambios mínimos en las condiciones iniciales, haciendo que las predicciones a largo plazo sobre la evolución del sistema sean imposibles. Para que pueda presentarse este comportamiento el sistema dinámico debe depender de al menos tres variables (o grados de libertad) diferentes explícitamente si es un sistema continuo, o de dos variables si es un mapa discreto (Strogatz 2000).

2.5 Cantidades Invariantes

Un invariante dinámico (o simplemente invariante) es una cantidad que describe la naturaleza dinámica de un sistema, con la propiedad de ser independiente de las coordenadas del sistema (Small 2005). Esta es la definición de Invariante dinámico al menos desde el punto de vista de los sistemas lineales, sin embargo para los sistemas no lineales esta definición es un poco diferente. Debido a la naturaleza de los sistemas no lineales en vez de hablar de invariabilidad se debe hablar de “poca sensibilidad” ante transformaciones suaves del espacio de fase (Kantz y Schreiber 2006).

Los invariantes dinámicos se tratarán ampliamente en este trabajo, ya que el cálculo de estas cantidades permitirá revelar características importantes de las series de tiempo. Algunos de estos invariantes son los *Exponentes de Lyapunov*, la *Dimensión de Correlación*, y *Entropía de la Información*.

En este capítulo se desarrolló una descripción de los conceptos básicos que serán necesarios para la correcta comprensión del trabajo, los cuales hacen parte de la teoría de sistemas dinámicos no lineales. En el capítulo siguiente se describirán las herramientas utilizadas para los análisis de no linealidad en series de tiempo.

3. Herramientas de análisis

En este capítulo se presentarán las herramientas que serán utilizadas para el análisis de series de tiempo. Las herramientas que se verán a continuación son el exponente de Hurst, utilizado para hacer una inspección inicial de la serie; el rezago óptimo de embebimiento y la dimensión de embebimiento, utilizados para hacer una emulación del espacio de fase del sistema dinámico que genera la serie de tiempo; y por último, el modelo de pronóstico lineal constante, el cual es una herramienta muy sencilla de pronóstico, con la cual se pretende rectificar la predictibilidad del sistema a partir del embebimiento construido.

En general el análisis de datos producto de la evolución temporal de sistemas caóticos, es considerado más experimental que teórico (Sprott 2003) y menos desarrollado que la teoría de sistemas dinámicos caóticos. Prueba de ello es la existencia de diversas formas para calcular las herramientas expuestas en este capítulo. En esta investigación se expondrán y utilizarán algunas de las metodologías más ampliamente aceptadas.

3.1 Análisis de re-escalamiento y el exponente de Hurst

Esta metodología fue desarrollada por el hidrólogo inglés Harold Edwin Hurst (Peters 1994), el cual participó en el proyecto de construcción de la presa del río Nilo, y su estudio se basó extensamente en modelar el caudal del mismo.

En la construcción de presas, los hidrólogos se interesan particularmente en la determinación tanto del caudal de entrada, el cual puede verse afectado por diferentes factores climatológicos; como el de salida, afectado por el consumo de agua. En general la suposición más común es que el flujo de entrada a un embalse se comporta como un proceso aleatorio, sin embargo Hurst demostró algo diferente.

Tras estudiar la información de alrededor de 847 años que los Egipcios guardaron sobre desbordamientos del río Nilo, Hurst descubrió la existencia de ciclos aperiódicos (Peters 1994) y llegó a la conclusión de que el comportamiento del nivel del río no era aleatorio. Sin embargo siguiendo los procedimientos estadísticos tradicionales, se encontró con que las correlaciones entre las observaciones no eran significativas, así que decidió desarrollar su propia metodología de análisis.

3.1.1 El método

(Peters 1994) presenta una descripción detallada del algoritmo utilizado para el cálculo del exponente de Hurst. Esta cantidad se encuentra estrechamente ligado con el trabajo de Einstein sobre el movimiento Browniano, el cual relaciona la distancia R que puede recorrer una partícula suspendida en un fluido, con el tiempo T que le toma recorrer dicha distancia. Esta relación está determinada mediante la ecuación (3.1)

$$R = T^{0.5} \quad (3.1)$$

La ecuación (3.1) es ampliamente conocida en finanzas y se utiliza regularmente para anualizar la volatilidad o la desviación estándar. Hurst pensó que podría basarse en ella para chequear la aleatoriedad del comportamiento de los desbordamientos en el río Nilo e ideó la siguiente metodología (Peters 1994).

Sea $x = x_1, x_2, \dots, x_n$ una serie temporal de n componentes de la variable a estudiar. El valor medio de esta serie temporal se define como

$$x_m = (x_1 + x_2 + \dots + x_n)/n \quad (3.2)$$

Y la desviación estándar como

$$s_n = n^{-1/2} * \sqrt{\sum (x_r - x_m)^2} \quad (3.3)$$

Se sustrae la media de la serie como muestra la ecuación (3.4)

$$Z_r = (x_r - x_m) \quad ; \quad r = 1, \dots, n \quad (3.4)$$

Y se genera una serie de sumas acumuladas como muestra la ecuación (3.5)

$$Y_r = \sum_{i=1}^r z_i \quad ; \quad r = 1, \dots, n \quad (3.5)$$

Nótese que la última componente del vector Y_r debe ser cero, debido a la estandarización realizada en la ecuación (3.4). Por último se calcula el rango ajustado de x denominado R_n como se muestra en la ecuación (3.6). Puede demostrarse que el valor R_n es no negativo

$$R_n = \max(Y_1, \dots, Y_n) - \min(Y_1, \dots, Y_n) \quad (3.6)$$

El rango ajustado R_n representa la distancia máxima que se mueve el proceso en un tiempo dado n . Luego, podría pensarse en aplicar la ecuación (3.1) tomando $T = n$, sin embargo esta solo aplica para movimientos Brownianos, es decir, series temporales con media cero y desviación estándar 1. Hurst encontró que una generalización de la ecuación (3.1) puede escribirse como se muestra en la ecuación (3.7)

$$(R/S)_n = c * n^H \quad (3.7)$$

Donde el valor $(R/S)_n$ es denominado el rango reescalado de x y c es una constante. Este rango reescalado está expresado en términos de la desviación estándar local. En general el valor del rango reescalado cambia con el incremento de n por medio de una ley exponencial dependiente del parámetro H , el cual se denomina el Exponente de Hurst. Luego calculando el logaritmo de la ecuación (3.7) se tiene

$$\text{Log}(R/S)_n = \text{Log}(c) + H * \text{Log}(n) \quad (3.8)$$

Por lo tanto el exponente de Hurst H puede ser calculado a partir del ajuste lineal sobre la grafica del logaritmo del rango reescalado $(R/S)_n$ versus el logaritmo del tiempo n (ecuación (3.8)), este valor será un número entre 0 y 1 dependiendo de las características de la serie estudiada.

3.1.2 Interpretación

Cuando este valor cumple la condición $0 < H < 0.5$ se dice que la serie presenta un comportamiento anti-persistente. En este caso se tiene que la serie de tiempo recorre un rango más corto (se aleja menos de su valor inicial) en un intervalo de tiempo determinado en comparación al rango recorrido por un proceso aleatorio, es decir, presenta un comportamiento similar al de reversión a la media. Sin embargo no puede asegurarse que la media permanezca estable aunque debe presentar movimientos erráticos que confinan su variación a un rango más estrecho que el generado por un proceso aleatorio.

Para el caso en que $H = 0.5$ se dice que la serie se comporta como un proceso aleatorio, o en general independiente. Se debe tener en cuenta que este resultado no es exclusivo de una distribución Gaussiana, por lo cual es posible que los datos sean generados aleatoriamente con una distribución uniforme, chi-cuadrada o t-student, por dar algunos ejemplos.

Y por último, cuando $0.5 < H < 1$ se dice que la serie presenta un comportamiento persistente, lo cual indica que la serie de tiempo recorre un rango más amplio que un proceso aleatorio. Este tipo de comportamientos se presenta en procesos con efectos de memoria larga la cual es independiente de la escala de tiempo.

Se espera que los datos resultantes de un sistema con comportamiento caótico presenten un exponente de Hurst entre $0.5 < H < 1$, es decir, que la serie sea persistente, ya que la alta sensibilidad al cambio en las condiciones iniciales¹ produce que la serie de tiempo exhiba efectos de memoria larga, por lo cual si dos sistemas se inician en valores infinitesimalmente cercanos, tendrá una evolución completamente diferente a largo plazo. En conclusión, esta es la primera pista que puede sugerir la presencia de determinismo en los datos analizados: un exponente de Hurst entre $0.5 < H < 1$.

Aunque como se dijo anteriormente esta es una primera pista de la presencia de determinismo en la serie, no es una causa suficiente para afirmar que hay presencia de determinismo y menos aún de la presencia de caos. Podría decirse que en realidad lo mejor que se puede tener es una sospecha y en base a esta sospecha utilizar otras herramientas de análisis que permitan obtener conclusiones al respecto.

¹ En la sección de conceptos básicos se mencionó que esta es una de las características de los sistemas con dinámica caótica. Cuando se introduzca el concepto de Exponentes de Lyapunov se explicará mejor esta afirmación.

Adicionalmente al algoritmo descrito en (Peters 1994), se presentan dos medidas interesantes, una que determina la confianza sobre el resultado de la estimación de H que es el cálculo del valor esperado del análisis de reescalamiento $E(R/S)$; y el otro llamado el estadístico V (V -Statistic), el cual es útil para detectar ciclos aperiódicos que pueda tener la serie de tiempo. El cálculo de $E(R/S)$ y el estadístico V proporcionarán una mayor confiabilidad al resultado de H . Por otro lado mientras más alta sea la cantidad de datos utilizada para la estimación, más alta será también la confiabilidad en los valores H , $E(R/S)$ y V .

El valor esperado del análisis de reescalamiento $E(R/S)$ se define como el valor que debería tener el exponente de la serie analizada en caso de ser un movimiento browniano geométrico. El valor de $E(R/S)$ se calcula por medio de la fórmula

$$E(R/S)_n = \left(\frac{(n-0.5)}{n}\right) \left(\frac{n\pi}{2}\right)^{-\frac{1}{2}} \sum_{r=1}^{n-1} \sqrt{\frac{n-r}{r}} \quad (3.9)$$

Donde n es el número de intervalos en los cuales se divide la muestra para el análisis de reescalamiento. La Figura 2 muestra un ejemplo del cálculo del exponente de Hurst para el sistema caótico de Chua para una serie de **1500** muestras.

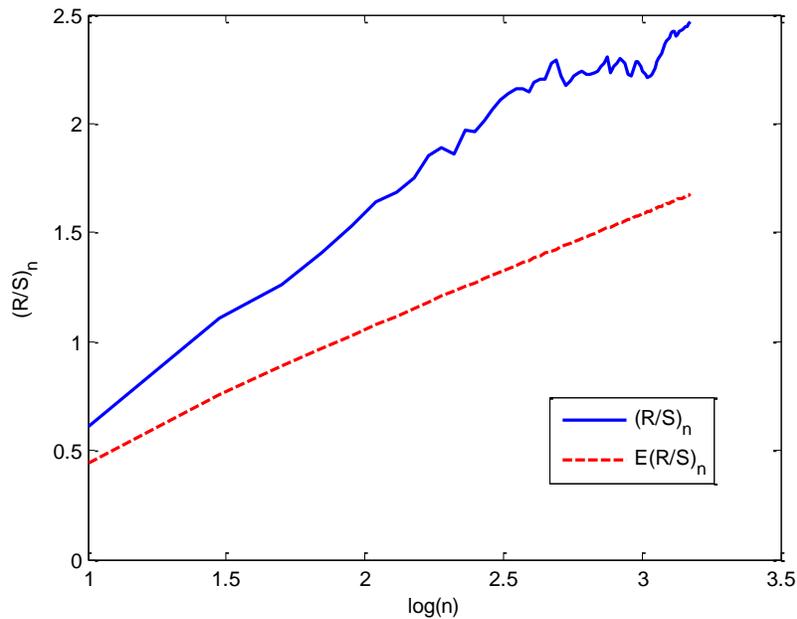


Figura 2: Grafica del análisis de reescalamiento del sistema dinámico de Chua cuando el sistema está en régimen caótico, claramente se muestra como su pendiente es mayor que la del valor esperado. El valor de $H = 0.7305$ mientras que el valor esperado es $E(R/S) = 0.5189$

La varianza se calcula de la siguiente forma

$$Var(H) = \frac{1}{T} \quad (3.10)$$

Donde T es el número total de datos. Luego la medida de confianza consiste en calcular el número de desviaciones estándar entre el valor esperado $E(R/S)$ y el valor calculado de H y determinar si la diferencia es significativamente amplia. Si el valor de $H > E(R/S) + 1,96\sigma$ podemos decir que la serie es persistente con un 95% de confianza.

Por otro lado el estadístico V se define de la siguiente manera:

$$V_n = (R/S)_n / \sqrt{n} \quad (3.11)$$

Al graficar esta medida resultará una línea horizontal en caso de que el estadístico R/S siga la ley de escalamiento descrita por la ecuación (3.1), es decir, sea una marcha aleatoria. En otras palabras, una grafica de V contra $\log(n)$ será horizontal si el proceso es un proceso aleatorio independiente. En caso de que el proceso sea persistente, el estadístico R/S evoluciona con una velocidad superior que la raíz cuadrada del tiempo ($H > 0.5$), luego la grafica presentará una pendiente ascendente. Por el contrario, si el proceso es antipersistente ($H < 0.5$), la grafica tendrá una pendiente descendente.

Una característica del estadístico V es que en los puntos donde se presenta un cambio de signo en la pendiente del estadístico se establece una longitud de periodo promedio, es decir, es posible establecer longitudes promedio de ciclo de una serie a partir del cálculo del estadístico V . En la Figura 3 se muestra una gráfica del estadístico V calculado para el sistema caótico de Rössler.

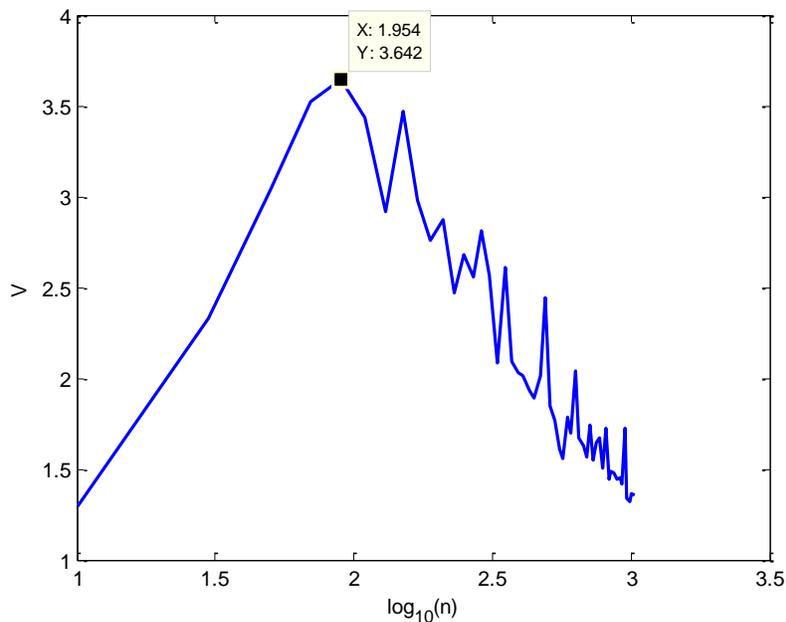


Figura 3: Estadístico V calculado para el sistema caótico de Rössler. Según el cambio de signo en la pendiente el ciclo promedio de la serie ocurre aproximadamente cada $10^{1.954} \approx 90$ muestras.

Cabe anotar que, según (Peters 1994), las series con presencia de dinámica lineal y de memoria corta como, por ejemplo, con presencia de procesos *AR*, *MA*, *ARMA*, *ARIMA*, pueden sesgar el resultado del análisis de re-escalamiento mostrando un fenómeno de persistencia.

Debe tenerse en cuenta que el objetivo principal del desarrollo del análisis de re-escalamiento, es detectar persistencia mediante un fenómeno de memoria larga en la serie, típico de las series con dinámica no lineal e incluso caos (Peters 1994) (Díaz 2000). El sesgo que se genera en el resultado puede ser reducido, según (Peters 1994), mediante el un procedimiento llamado prewhitening ó detrending, el cual consiste en la aplicación de un filtrado *AR(1)* sobre la serie a analizar. De esta forma se espera eliminar gran parte del sesgo generado por estos procesos.

En el caso en que el residuo del filtrado de la serie sea un proceso aleatorio ó antipersistente, quiere decir que la totalidad de la dinámica era lineal y/o con memoria únicamente de corto plazo, lo cual es perfectamente modelable mediante las estructuras *AR*, *MA*, *ARMA* y *ARIMA*.

3.2 Teoría de embebimiento

Esta teoría propone un cambio de paradigma al estudio que actualmente se le da a las series de tiempo en Colombia, tanto desde el campo de la estadística como desde los métodos heurísticos y la inteligencia artificial. Las técnicas que se mostrarán a continuación suponen que el conjunto de datos observados en una serie de tiempo tienen su origen en un sistema dinámico determinista con una dimensión desconocida.

Dependiendo de la complejidad que presenten los datos de la serie puede resultar más conveniente modelarlos como un sistema determinista (cuando los análisis llevan a pensar que los datos provienen de un sistema dinámico de baja dimensión) ó modelarlos como un proceso estocástico (cuando los análisis llevan a pensar que los datos provienen de un sistema con dimensión muy alta) (Strogatz 2000).

Luego, si se supone que el sistema puede ser descrito por medio de un conjunto reducido de variables debe aprovecharse la información disponible en la serie de tiempo de la mejor forma posible. De este modo según (Abarbanel 1996) si se muestrea una señal $s(t)$ en intervalos de tiempo τ_s iniciando en un punto t_0 , la serie de tiempo será de la forma $s(n) = s(t_0 + n\tau_s)$, y la evolución observada es equivalente a un mapa discreto generando $s(k + 1)$ a partir de $s(k)$.

Luego, se puede representar el siguiente flujo

$$\frac{dx(t)}{dt} = F(x(t)) \quad (3.12)$$

Como una evolución muestreada de forma finita

$$x(t_0 + (n + 1)\tau_s) \approx x(t_0 + n\tau_s) + \tau_s F(x(t_0 + n\tau_s)) \quad (3.13)$$

De esta forma las observaciones tendrán el siguiente comportamiento

$$s(t_0 + k\tau_s) \rightarrow s(t_0 + (k + 1)\tau_s) \quad (3.14)$$

$$s(k) \rightarrow s(k + 1)$$

Que es en realidad la serie temporal del fenómeno que se quiere describir. Ahora la pregunta más conveniente en este punto es ¿cómo se puede obtener una representación para el espacio de fase de $F(x(t))$ a partir de los datos observados $s(t)$? La respuesta a esta pregunta puede encontrarse en artículos como el de (Eckmann y Ruelle 1985), donde se presenta el teorema atribuido a Takens y Mañé, descrito a continuación.

3.2.1 Teorema de Takens.

Suponga un sistema dinámico $x(n) \rightarrow F(x(n)) = x(n + 1)$ donde el espacio de fase de $x(t)$ es multidimensional. Si se puede observar un valor escalar $h(\blacksquare)$, de alguna función vectorial de las variables dinámicas $g(x(n))$, entonces la estructura geométrica de la dinámica multivariable puede ser “desplegada” del conjunto de medidas escalares $h(g(x(n)))$ en un espacio construido a partir de nuevos vectores cuyas componentes consisten en $h(\blacksquare)$ aplicada a las potencias de $g(x(n))$. De esta forma (3.15) describe el movimiento en un espacio euclidiano d-dimensional.

$$y(n) = \left[h(x(n)), h(g^{T_1}(x(n))), h(g^{T_2}(x(n))), \dots, h(g^{T_{d-1}}(x(n))) \right] \quad (3.15)$$

Con algunas condiciones generales de suavidad en las funciones $h(\blacksquare)$ y $g(x)$ se tiene que para un d lo suficientemente grande varias propiedades importantes de la señal desconocida $x(n)$ son reproducidas sin ambigüedad en el nuevo espacio vectorial $y(n)$, más aun, $y(n)$ será un homeomorfismo de $x(n)$.

La implementación de este teorema exige escoger convenientemente las funciones $h(\blacksquare)$ y $g(x)$, así que se seleccionarán de forma que sea fácil utilizar los datos observados. De esta forma se tomará $h(x(n))$ como la serie de tiempo de observaciones $s(n)$.

$$h(x(n)) = s(n) \quad (3.16)$$

y la función $g(x)$ se seleccionará como el operador rezago aplicado al vector x con un periodo de muestreo τ_s , y la potencia T_k de $g(x)$ será entonces

$$g^{T_k}(x) = x(n + T_k) = x(t_0 + (n + T_k)\tau_s) \quad (3.17)$$

Y así las componentes de $y(n)$ serán

$$y(n) = [s(n), s(n + T_1), s(n + T_2), \dots, s(n + T_{d-1})] \quad (3.18)$$

Si se selecciona $T_k = kT$, es decir que los múltiplos enteros de un rezago común, $y(t)$ queda de la siguiente manera

$$y(n) = [s(n), s(n + T), s(n + 2T), \dots, s(n + (d - 1)T)] \quad (3.19)$$

Luego las entradas escalares de la serie $s(n)$ será reemplazadas por los vectores de $y(n)$ en un espacio euclidiano d -dimensional y este capturará sin pérdida de información los aspectos invariantes de $x(n)$. Esto significa que se puede estudiar el sistema original a partir el espacio reconstruido y aprender esencialmente lo mismo que si se estuviera trabajando sobre el sistema “verdadero”.

El problema central de la reconstrucción es la selección de los parámetros: el rezago óptimo $T = \tau_s$, de forma que τ_s sea un múltiplo entero del periodo de muestreo $\tau_s = T\tau_s$; y el número mínimo de rezagos que deben usarse, llamado la dimensión de embebimiento² $d = d_s$, que debido al teorema de unicidad debe ser suficientemente grande para que las trayectorias no se traslapen (Eckmann y Ruelle 1985). El teorema de embebimiento de Takens no especifica en ningún momento como determinar el valor de estos parámetros, sin embargo (Broomhead y King 1985) hace algunas precisiones sobre el valor de la dimensión d_s y (Abarbanel 1996) recomienda algunas soluciones para el cálculo de τ_s . Estas precisiones serán mostradas a continuación.

Se supone que las medidas $s(n)$ son una parametrización temporal de la proyección unidimensional del atractor que gobierna el sistema dinámico a estudiar, y por ser una proyección se espera que las trayectorias del atractor se traslapen en dimensiones menores a la dimensión original. Los traslapos irán desapareciendo a medida que se incremente el número de vectores en $y(n)$, es decir, a medida que la dimensión d_s se incremente de valor. El teorema especifica que si el sistema original se encuentra determinado por una dimensión d_A , escogiendo $d_s > 2d_A$ es suficiente para eliminar cualquier traslapo de la trayectoria, eliminando cualquier ambigüedad en la reconstrucción. A continuación se profundizará más en las técnicas para el cálculo de los parámetros τ_s y d_s .

3.2.2 Selección del tiempo de rezago.

El teorema de embebimiento de Takens no habla acerca de la elección del tiempo de rezago para la construcción del espacio d dimensional del sistema. Además, este permite cualquier tiempo de rezago, en la medida que se tenga una cantidad infinita de datos. Dado que en ningún caso real se tiene tal cantidad de puntos, es necesario buscar métodos para encontrar un valor confiable de τ_s .

Con base en la experiencia, particularmente con la relacionada con modelos de regresión y pronóstico, se puede esperar que una buena elección de un tiempo de rezago cumpla con:

² El término original en inglés es *embedding*, algunos autores lo traducen como *inserción*.

- Ser múltiplo del tiempo de muestreo de la serie, dado que sólo se tienen datos para estos tiempos. Un esquema de interpolación para “obtener más datos” es tan incierto como una estimación de la derivada de la serie.
- Si el tiempo de rezago es muy corto las coordenadas $s(n)$ y $s(n + \tau_e)$ que serán usadas en el vector de reconstrucción de datos no serán lo suficientemente independientes. Es decir, no ha pasado suficiente tiempo para que el sistema evolucione en el espacio de estados.
- Si τ_e es muy grande cualquier relación entre $s(n)$ y $s(n + \tau_e)$ es insignificante y se pierde numéricamente.

De esta forma, se busca un tiempo de rezago que sea lo suficientemente grande para que $s(n)$ y $s(n + \tau_e)$ sean en cierta medida independientes pero no tanto como para que se pierda toda dependencia estadística. Para esto se procede a elegir τ_e basados en cualquiera de los métodos descritos a continuación.

Método basado en la función de autocorrelación.

Una forma de elegir el tiempo de rezago se basa en el cálculo de la función de autocorrelación muestral. Esta se define para una serie de tiempo escalar x_n de N muestras como:

$$\rho(\tau) = \frac{\sum_{n=1}^N (x_{n+\tau} - \bar{x})(x_n - \bar{x})}{\sum_{n=1}^N (x_n - \bar{x})^2} \quad (3.20)$$

Donde \bar{x} denota la media muestral de la serie x_n . El criterio para la selección de τ_e es elegir el tiempo de rezago donde $\rho(\tau)$ alcance su primer cero. De esta forma x_n y $x_{n+\tau_e}$ estarán linealmente incorrelacionados. En la literatura (Luo 2004), se ha encontrado que este criterio no funciona en todas las situaciones, debido a que los procesos con memoria larga tardan mucho tiempo en alcanzar el valor de cero para $\rho(\tau)$, lo cual hace impráctico su uso debido a que las muestras de datos son finitas. Por lo tanto se sugiere un criterio alternativo el cual consiste en elegir el tiempo de rezago τ_e donde se alcance un valor de $1/e$. Este método es más robusto para elegir un rezago adecuado, sin embargo no existe evidencia para demostrar que $1/e$ es un factor universal del criterio de la función de autocorrelación para la elección de rezagos adecuados.

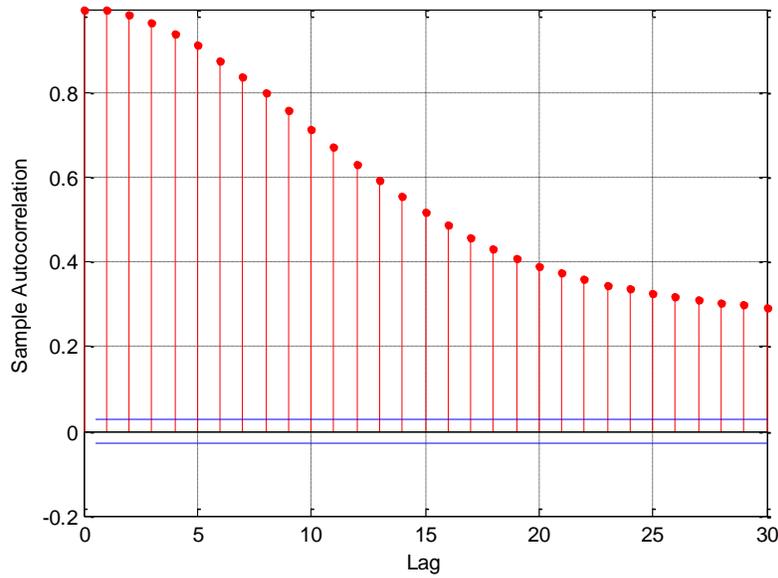


Figura 4: Función de autocorrelación para el sistema dinámico de Lorenz en régimen caótico, en este caso no es posible determinar el rezago óptimo τ_g .

Método basado en la función de medición de la información mutua.

La definición de la función de autocorrelación, presenta una medida lineal de la correlación entre los subconjuntos x_i y $x_{i+\tau_g}$, de esta forma si la serie se supone como la respuesta temporal de un sistema no lineal, existirá una pérdida de información al usar la función $\rho(\tau)$ para detectar algún tipo de correlación no lineal entre x_i y $x_{i+\tau_g}$. Bajo esta consideración, se propone el uso de un importante estadístico para medir la correlación no lineal llamado *información mutua*, cuya definición se muestra a continuación.

En teoría de probabilidad y teoría de la Información, la información mutua $I(X;Y)$ entre dos variables aleatorias es una cantidad que mide la dependencia mutua entre las dos variables.

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p_1(x)p_2(y)} \right) \quad (3.21)$$

Donde $p(x,y)$ es la función de distribución de probabilidad conjunta entre X y Y y $p_1(x)$ y $p_2(y)$ son las funciones de distribución marginales de X y Y respectivamente.

En el caso continuo la sumatoria se reemplaza con una integral doble definida:

$$I(X;Y) = \int_Y \int_X p(x,y) \log \left(\frac{p(x,y)}{p_1(x)p_2(y)} \right) dx dy \quad (3.22)$$

Donde $p(x,y)$ es la función de densidad de probabilidad conjunta de X y Y , y $p_1(x)$ y $p_2(y)$ son las funciones de densidad marginales de X y Y respectivamente. Estas definiciones son ambiguas debido a que la base de la función logaritmo no está definida. Para eliminar esta ambigüedad la función de información puede ser parametrizada como $I(X,Y,b)$ donde b es la base. Alternativamente, dado que la unidad de medida de la información mutua es el bit, puede asignarse la base como $b = 2$.

Intuitivamente, la información mutua mide la información que X y Y comparten, es decir, mide como conociendo una de estas variables se reduce la incertidumbre sobre la otra. Por ejemplo, si X y Y son independientes, entonces conocer X no da ninguna información sobre Y y viceversa, luego su información mutua es cero. Por otro lado, si X y Y son idénticas entonces toda la información contenida en X es compartida con Y , es decir, conociendo X se determina el valor de Y y viceversa. Como resultado, la información mutua es igual a la incertidumbre contenida en Y (o X) sola, esto es, la entropía de Y (o X).

La información mutua cuantifica la dependencia entre las distribuciones de X y Y . Luego, en caso de que X y Y fueran variables independientes $I(X,Y) = 0$. Esto es fácil de observar: si X y Y son independientes entonces $p(x,y) = p(x)p(y)$ y por tanto:

$$\log\left(\frac{p(x,y)}{p_1(x)p_2(y)}\right) = \log 1 = 0 \quad (3.23)$$

Más aún, la información Mutua es una función semidefinida positiva ($I(X;Y) \geq 0$) y simétrica ($I(X;Y) = I(Y;X)$).

Para elegir el tiempo de rezago, el criterio que se propone es tomar el primer mínimo de la función de información mutua calculado entre las componentes de la misma serie $I(x_n; x_n)$. Este criterio presenta grandes ventajas debido a que la información mutua es la herramienta correcta para analizar series de tiempo en sentido no lineal (Abarbanel 1996). Sin embargo se presentan algunos inconvenientes, el más significativo es que para algunas series la función decrece de forma monótona y no presenta mínimos. Este inconveniente puede presentarse cuando la serie analizada es persistente pero no estacionaria, es decir, tiene componentes de tendencia muy marcadas. En estos casos el método no entrega un valor confiable del rezago óptimo, y se debe optar por cambiar de metodología ó trabajar con una modificación de la serie. Una modificación de la serie puede consistir en modelar las dinámicas de los incrementos de $x(n)$ en lugar de trabajar sobre la serie (Small 2005).

Un ejemplo simple de la modificación de la serie podría ser asumir un proceso recursivo así

$$x(n+1) = x(n) + f(x(n)) \quad (3.24)$$

Y la dinámica a analizar pasa de ser la de $x(n + 1)$ a la de $f(x(n))$. De esta misma forma puede proponerse trabajar con incrementos logarítmicos de la serie $x(n)$, o incluso con la serie eliminando la tendencia. Un ejemplo de la función de información mutua se presenta en la Figura 5.

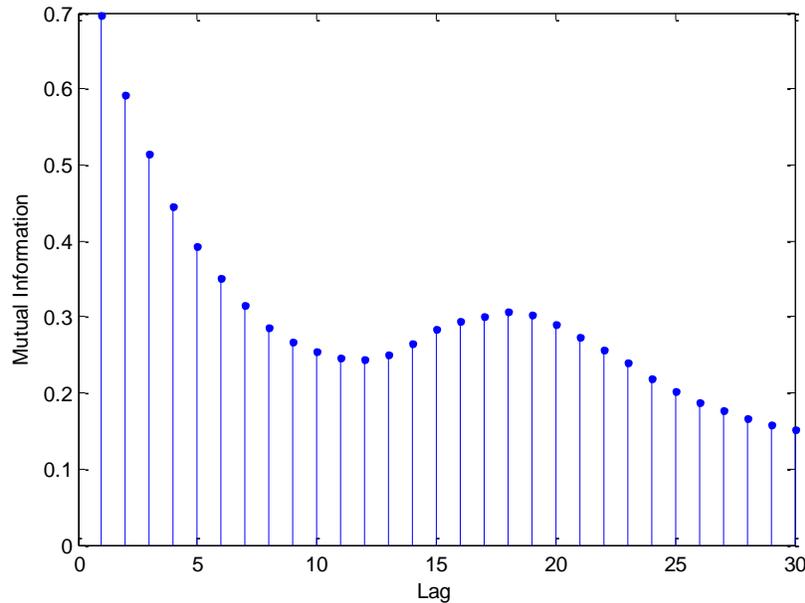


Figura 5: Información mutua para el sistema dinámico de Lorenz en régimen caótico, en este caso el rezago óptimo claramente es $\tau_g = 12$

Método del Periodo aproximado.

La idea de los dos métodos anteriores es elegir un rezago en el cual x_i y $x_{i+\tau_g}$ estén razonablemente incorrelacionados pero que aún estén cerca uno de otro. Cuando los datos muestran una periodicidad fuerte, un valor de τ_g igual a un cuarto de la longitud del periodo principal generalmente da un buen embebimiento. Este rezago es aproximadamente el mismo del primer cero de la función de autocorrelación. Con esta asignación se supone que el sistema dinámico subyacente se puede embeber en un espacio con dimensión $d_g \geq 4$ (la determinación de d_g se ilustrará en la siguiente sección), ya que para estas dimensiones el periodo estará completamente cubierto (Small 2005). Además es una forma muy sencilla de calcular el rezago.

3.2.3 Selección de la dimensión de embebimiento.

Al obtener el tiempo de rezago, se busca determinar la dimensión global entera donde se tienen el número necesario de coordenadas para desplegar las orbitas observadas y eliminar los traslapos producidos por la proyección del atractor en un espacio de dimensión menor. Para esto se analiza al conjunto de datos $y(n)$ y se observa cuando se presentan estos traslapos.

Se podría pensar que la mejor decisión es simplemente asignar una dimensión alta, sin embargo esto puede generar algunos percances como la necesidad de una mayor cantidad de datos para asegurar un mejor modelado de la trayectoria, además de que con cada unidad aumentada en la dimensión se aumenta la cantidad de ruido introducida (Sprott 2003). Es por esto que debe determinarse la menor dimensión que despliegue el atractor.

La dimensión más baja en la que se despliega el atractor sin que ocurran los traslapos en las orbitas se llama dimensión de embebimiento d_E , siendo d_E un número entero. Si toman dos muestras $s_A(n)$ y $s_B(n)$ del mismo sistema, no se garantiza que al calcular la dimensión d_E sea igual para cada una de las muestras. Cada medición con sus rezagos brinda una combinación no lineal diferente de las variables dinámicas originales y puede dar un mapeo global no lineal del verdadero espacio en el espacio reconstruido de dimensión d_E , donde la suavidad y la unicidad de las trayectorias se preservan.

El teorema de embebimiento dice que si la dimensión de un atractor definido por las orbitas es d_A , entonces puede desplegarse el atractor en un espacio de dimensión entera d_E donde $d_E > 2d_A$. Esta no es necesariamente la dimensión para desplegarlo, pero es suficiente y dice cuando detener la adición de componentes al vector de rezagos.

Método de los falsos vecinos cercanos

Suponga que se tiene una reconstrucción del espacio de estados de dimensión d con vectores de datos usando el tiempo de rezago sugerido por la función de información mutua como se muestra en (3.25)

$$y(n) = [s(n), s(n+T), s(n+2T), \dots, s(n+(d-1)T)] \quad (3.25)$$

Si se examina el vecino más cercano en el espacio de estados del vector $y(n)$ en el tiempo n este será un vector

$$y^{NN}(n) = [s^{NN}(n), s^{NN}(n+T), s^{NN}(n+2T), \dots, s^{NN}(n+(d-1)T)] \quad (3.26)$$

y su valor en el tiempo n tendrá una relación con el valor del vector $y(n)$ en ese mismo tiempo. Si el vector $y^{NN}(n)$ es realmente un vecino de $y(n)$, entonces proviene de una vecindad de $y(n)$ en sus orígenes dinámicos. El vector está adelante o atrás de $y(n)$ a través de la órbita si los pasos de tiempo a lo largo de la órbita son los suficientemente pequeños, ó llegan a la vecindad de $y(n)$ por evolución de la órbita a lo largo del atractor. Dado que los atractores para los sistemas deterministas son compactos, cada punto en el espacio de estados tendrá tantos vecinos como el número de datos que hacen que el espacio de estados quede bien poblado.

Si $y^{NN}(n)$ es un vecino falso de $y(n)$ habiendo llegado a su vecindario por una proyección de una dimensión más alta debido a que la dimensión actual d no despliega el atractor, entonces

aumentando la dimensión a $d + 1$ se moverá este vecino falso de $y(n)$. Inspeccionando todos los datos de la serie y verificando con cuál dimensión se remueven todos los vecinos falsos, se removerán secuencialmente los traslajos de las orbitas de dimensión más baja. Cuando no se detecten más falsos vecinos cercanos, se habrá identificado el valor d_g en el cual el atractor está completamente desplegado.

Ahora es necesario un criterio para determinar cuando un punto $y(n)$ y su vecino más cercano $y^{NN}(n)$ vistos desde la dimensión d están cerca o lejos en la dimensión $d + 1$. Cambiando el valor de la dimensión d a la dimensión $d + 1$ la componente adicional para el vector $y(n)$ es sólo $x(n + d\tau_g)$, y la componente adicional para el vector $y^{NN}(n)$ es $x^{NN}(n + d\tau_g)$, ambas conocidas. Comparando la distancia entre los vectores $y(n)$ y $y^{NN}(n)$ en la dimensión d con la distancia de estos mismos vectores en la dimensión $d + 1$, se puede establecer con facilidad cuales vecinos son falsos y cuáles no. Sólo se necesita comparar $|x(n + d\tau_g) - x^{NN}(n + d\tau_g)|$ con la distancia euclidiana $|y(n) - y^{NN}(n)|$ entre los vecinos cercanos en la dimensión d . Si la distancia adicional es grande comparada con la distancia en la dimensión d entre los vecinos más cercanos, se tiene un vecino falso. Si la distancia no es muy grande, se tiene un verdadero vecino.

El cuadrado de la distancia euclidiana entre los puntos vecinos más cercanos vistos en la dimensión d es

$$R_d^2(n) = \sum_{m=1}^d [x(n + (m-1)\tau_g) - x^{NN}(n + (m-1)\tau_g)]^2 \quad (3.27)$$

Mientras en dimensión $d + 1$ es

$$R_{d+1}^2(n) = \sum_{m=1}^{d+1} [x(n + (m-1)\tau_g) - x^{NN}(n + (m-1)\tau_g)]^2 \quad (3.28)$$

De donde

$$R_{d+1}^2(n) = R_d^2(n) + |x(n + d\tau_g) - x^{NN}(n + d\tau_g)|^2 \quad (3.29)$$

La distancia entre los puntos cuando son vistos en dimensión $d + 1$ relativa a la distancia en la dimensión d es

$$\sqrt{\frac{R_{d+1}^2(n) - R_d^2(n)}{R_d^2(n)}} = \frac{|x(n + d\tau_g) - x^{NN}(n + d\tau_g)|}{R_d(n)} \quad (3.30)$$

cuando esta cantidad es mayor que un umbral R_{tol} , se tiene un falso vecino. En la práctica el umbral R_{tol} para $\frac{|x(n + d\tau_g) - x^{NN}(n + d\tau_g)|}{R_d(n)}$ usado para definir un vecino falso es un número alrededor

de 15. Esto varía con el número de datos para conjuntos pequeños, pero en la medida en que se muestrean todas la regiones del atractor, la variación de los vecinos falsos con el número de datos es muy pequeña.

Método de Cao

En el trabajo de Cao (Cao 1997) se propone un método para determinar la dimensión mínima de embebimiento para una serie de tiempo escalar, este método es llamado usualmente Método de Cao o el Método de Falsos Vecinos Cercanos Promediados. Este procedimiento cuenta con las siguientes ventajas:

- No contiene ningún parámetro subjetivo excepto el tiempo de rezago τ_ε para el embebimiento.
- No depende significativamente de cuantos puntos están disponibles.
- Puede distinguir señales determinísticas de señales estocásticas.
- Trabaja correctamente con series que provienen de sistemas de alta dimensión.
- Es computacionalmente eficiente.

El método se basa en la construcción del vector de rezagos de la serie. El vector de dimensión d_ε se define como:

$$y(n) = [s(n), s(n + T), s(n + 2T), \dots, s(n - (d_\varepsilon - 1)\tau_\varepsilon)] \quad (3.31)$$

Con $n = 1, \dots, N - (d_\varepsilon - 1)\tau_\varepsilon$. Similarmente al método de los Falsos Vecinos Cercanos, en este método se define la cantidad:

$$a(n, d) = \frac{\|y_n(d+1) - y_{k(n,d)}(d+1)\|}{\|y_n(d) - y_{k(n,d)}(d)\|} \quad (3.32)$$

Donde $\|\bullet\|$ es la norma del máximo, que para este caso se usa de la forma:

$$\|y_k(m) - y_l(m)\| = \max_{0 \leq j \leq m-1} |x(k + j\tau_\varepsilon) - x(l + j\tau_\varepsilon)| \quad (3.33)$$

Los resultados del método son iguales independientemente de la elección de la norma para medir la distancia entre los vectores $y_n(d)$. El número $k(n, d)$ es un entero tal que el vector de dimensión d , $y_{k(n,d)}(d)$ es el vecino más cercano del vector $y_n(d)$. La ventaja del Método de Falsos Vecinos Cercanos Promediados radica en definir la cantidad promediada:

$$E(d) = \frac{1}{N - d\tau_\varepsilon} \sum_{n=1}^{N - d\tau_\varepsilon} a(n, d) \quad (3.34)$$

Y definir posteriormente el parámetro:

$$E_1(d) = \frac{E(d+1)}{E(d)} \quad (3.35)$$

En la medida que d crece, el valor del parámetro $E_1(d)$ se estabiliza en un valor d_0 cuando la serie proviene de un proceso determinístico, en este caso se espera que d_0 se aproxime a 1. Cuando esto sucede la mínima dimensión de embebimiento será $d_e = f^{-1}(d_0 + 1)$, es decir, el dominio del valor en el cual $d_0 \approx 1$ (Cao 1997).

En la figura 6 se muestra un ejemplo del método de Cao para el cálculo de la dimensión de embebimiento. Nótese que para los valores mayores que $d_e = 8$ el valor de $E_1(d)$ también debe permanecer cerca a 1.

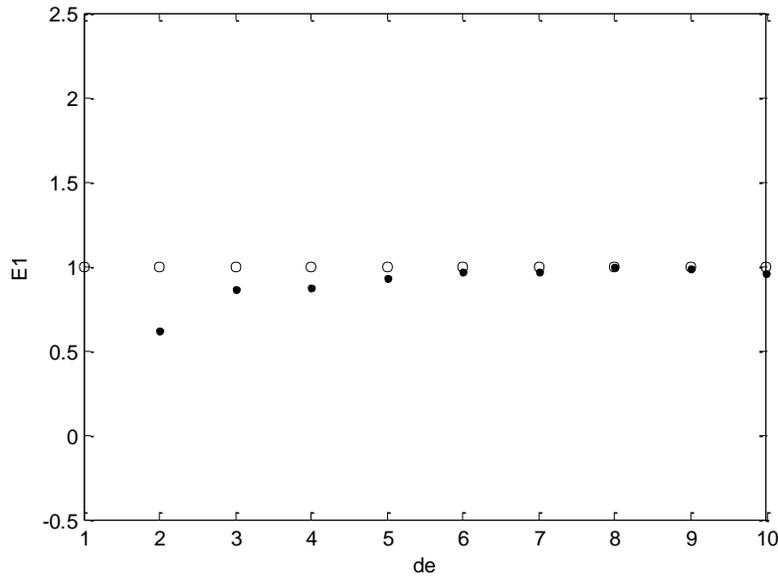


Figura 6: Dimensión de embebimiento del sistema dinámico de Chua en régimen caótico, el algoritmo entrega un valor $d_e = 8$.

3.3 Método de de pronóstico

El método seleccionado para la elaboración del pronóstico es de hecho bastante simple. Después de encontrar parámetros para el embebimiento se construye una matriz de rezagos D a partir de la serie original $\{y_n\}$, donde el número de filas es d_e y el número de columnas viene dado por $N - (d_e - 1)\tau_e$, donde N es el número de muestras de la serie de tiempo.

$$D = \begin{bmatrix} y_{(d_e-1)\tau_e} & \cdots & y_N \\ \vdots & & \vdots \\ y_1 & \cdots & y_{N-(d_e-1)\tau_e} \end{bmatrix} \quad (3.36)$$

Cada fila de la matriz D equivale a una dimensión del espacio de fase reconstruido y cada columna corresponde a un punto en dicho espacio.

La estructura propuesta por (Small 2005) es un modelo de pronóstico local constante (LCM) el cual es recursivo y se construye de la siguiente manera. Se selecciona el punto x_f que equivale a el punto d_ε -dimensional más reciente de los datos disponibles en la matriz D , luego se busca un número m determinado de vecinos cercanos $\{x_N\}$ al punto seleccionado, con los cuales puede formarse la matriz D_N .

El modelo LCM utiliza la matriz D_N en una regresión lineal por mínimos cuadrados para ajustar el valor de los parámetros de un vector λ contra el vector final de la matriz de datos disponibles x_f por medio del uso de una regresión múltiple.

Si se agrupan los puntos siguientes a los m vecinos cercanos $\{x_N\}$ de x_f , denotados como $\{x_{N+1}\}$ se puede construir la matriz D_{N+1} , la cual al ser multiplicada por el vector de parámetros m produce la estimación para el primer punto d_ε - dimensional por fuera de la muestra, es decir

$$D_{N+1} \cdot \lambda = \hat{x}_{f+1} \quad (3.36)$$

Donde la única componente relevante del ajuste es el valor de la primera componente \hat{y}_{n+1} del vector \hat{x}_{f+1} ya que las demás componentes son conocidas debido a la construcción en rezagos de la matriz D . De esta forma punto d_ε - dimensional pronosticado en el espacio de fase sería en realidad:

$$\hat{x}_{f+1} = \begin{pmatrix} \hat{y}_{n+1} \\ y_{n-\tau_\varepsilon+1} \\ \vdots \\ y_{n-(d_\varepsilon-1)\tau_\varepsilon+1} \end{pmatrix} \quad (3.36)$$

Luego el error de pronóstico viene dado por la diferencia entre el valor estimado \hat{y}_{n+1} y el valor real y_{n+1} . Al obtenerse el valor real y_{n+1} , se añade el punto x_{f+1} al final de la matriz D y se comienza el proceso de pronóstico nuevamente.

En este capítulo se describió la primera prueba que permite tener una sospecha inicial sobre la presencia ó ausencia de caos determinista en series de tiempo. El cálculo del exponente de Hurst determina la existencia de persistencia en series de tiempo, sin embargo no es una medida concluyente sobre la presencia de caos determinista, es por esto que en capítulos posteriores se introducirán algunas pruebas adicionales para soportar las conclusiones a las que se llegue.

Por otro lado se mostró cómo es posible construir un espacio de fase a partir del cálculo de una dimensión de embebimiento y un rezago óptimo. La construcción de un espacio de fase es necesaria para algunas de las herramientas que se expondrán en capítulos posteriores. Adicionalmente se presentó el modelo local de pronóstico constante (LCM), que más que ser un método robusto de pronóstico, se evalúa para chequear si es suficiente la información capturada

por la reconstrucción del espacio de fase. Este será probado para evaluar la reconstrucción del espacio de fase en el análisis de la TRM más adelante.

En el capítulo siguiente se mostrarán algunas medidas invariantes utilizadas para la caracterización de los atractores extraños de los sistemas dinámicos que presentan caos, a partir de estas medidas podrán desarrollarse las primeras conclusiones.

4. Invariantes Topológicos en Sistemas No lineales

Así como en los sistemas lineales se presentan cantidades que no cambian con el tiempo, llamadas invariantes, en los sistemas no-lineales podemos encontrar cantidades con un comportamiento similar. Sin embargo en el caso de los sistemas no lineales cuando nos referimos a “invariabilidad” estamos hablando de una “baja sensibilidad” a algún tipo de cambio.

En (Kantz y Schreiber 2006) se afirma que dos medidas son de gran relevancia en la caracterización del caos determinista: el máximo exponente de Lyapunov y la dimensión de correlación. Su relevancia radica en la invariabilidad de estas cantidades bajo transformaciones suaves del espacio de fase. Sin embargo, esto es cierto solo en un caso ideal, es decir, para series de tiempo infinitas y libres de ruido. En el caso que compete a este trabajo, un buen algoritmo aplicado a un conjunto de datos lo suficientemente grande y con un nivel relativamente bajo de ruido debería producir resultados robustos a pequeños cambios en los parámetros del algoritmo.

El máximo exponente de Lyapunov y la dimensión de correlación son solo dos miembros de la gran familia de invariantes que se han descubierto y son los más utilizados debido a que son los que presentan mejor comportamiento al ser extraídos de datos experimentales, sin embargo (Small 2005) argumenta que hay otra cantidad invariante que puede ser de gran utilidad como es la entropía de la información. A continuación se introducen tres invariantes que serán utilizadas para determinar la existencia de caos en series de tiempo. Estas cantidades son la Entropía de la información, los Exponentes de Lyapunov y la Dimensión de Correlación.

4.1 Entropía de la información

Una descripción general de Entropía desde el punto de vista termodinámico es la “medida del desorden de un sistema dinámico”. Shannon, durante sus estudios de codificación, transmisión y decodificación de información, introdujo una noción del grado de desorden ó incertidumbre al intentar predecir el resultado de un evento probabilístico. Supóngase que se tiene un experimento con r posibles resultados cuyas probabilidades son p_1, p_2, \dots, p_r . Luego la entropía de Shannon se define como:

$$H_s = \sum_{i=1}^r p_i \log\left(\frac{1}{p_i}\right) \quad (4.1)$$

En el caso en que una de las $p_i = 1$ y el resto cero, el resultado de $H_s = 0$ (definiendo $p \log(1/p) \equiv 0$ para $p = 0$), que es consistente con el hecho de que no haya incertidumbre en el resultado del experimento ya que se sabe con certeza que el evento p_i ocurrirá. La incertidumbre será máxima cuando todos los eventos son igualmente probables, es decir $p_1 = p_2 = \dots = p_r = 1/r$, en cuyo caso la entropía H_s toma su máximo valor probable de

$-\log(r)$. Nótese que en este caso H_s crece cuando aumenta r debido al aumento de la incertidumbre al tener más eventos con igual probabilidad (Ott, Sauer y Yorke 1994).

Kolmogorov utilizó una aplicación de la medida de la entropía de Shannon en la teoría ergódica, creando su propia medida de la siguiente manera: Sea μ una medida ergódica invariante para un mapa G . Sea R una región acotada del espacio de fase que contiene la medida μ , y supóngase que R se divide en un número finito de subregiones $R = R_1 \cup \dots \cup R_r$ ($R_i \neq \emptyset$; $R_i \cap R_j = \emptyset, \forall i \neq j$; $i, j = 1, \dots, r$). Luego, se puede definir una función de entropía para la partición $\{R_i\}$ de la siguiente manera

$$H(\{R_i\}) = \sum_{i=1}^r \mu(R_i) \ln(\mu(R_i))^{-1} \quad (4.2)$$

lo que provee la información promedio obtenida cuando se sabe que una órbita se encuentra en una de las subregiones R_i (en ausencia de cualquier otra información). Luego Kolmogorov considerando las pre-imágenes $G^{-1}(R_i)$ de R_i examinó las r^2 intersecciones resultantes

$$R_i \cap G^{-1}(R_j) \quad (4.3)$$

para cada par (i, j) con $1 \leq i \leq r, 1 \leq j \leq r$, formando $\{R_i^{(2)}\}, 1 \leq i \leq r_2$, la partición de las intersecciones sin repetición. Luego formó una tercera partición $\{R_i^{(3)}\}$ para las r_3 intersecciones

$$R_i \cap G^{-1}(R_j) \cap G^{-2}(R_k) \quad (4.4)$$

para $i, j, k = 1, 2, \dots, r$. Así se forman más particiones de forma similar y se considera el valor

$$K(\mu, \{R_i\}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\{R_i^{(n)}\}) \quad (4.5)$$

Dado que

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(\{R_i^{(n)}\}) = \lim_{n \rightarrow \infty} [H(\{R_i^{(n+1)}\}) - H(\{R_i^{(n)}\})] \quad (4.6)$$

se puede pensar en $K(\mu, \{R_i\})$ como la información promedio ganada al pasar de la partición n a la partición $n + 1$, para n grande. La entropía de Kolmogorov Sinai es por definición el supremo de $K(\mu, \{R_i\})$ sobre todas las posibles particiones iniciales.

$$K(\mu) \equiv \sup_{\{R_i\}} K(\mu, \{R_i\}) \quad (4.7)$$

Se puede entender K como la medida de la velocidad promedio a la cual se pierde predictibilidad en el sistema. Por otro lado, K es una medida positiva, así un sistema puramente aleatorio tendrá entropía infinita, mientras un sistema periódico tendrá entropía igual a cero (Sprott 2003). Debido a la sensibilidad a condiciones iniciales presentada por los sistemas caóticos, se dice que estos

generan información nueva (o pierde información debido a la disminución en la capacidad de predictibilidad) a medida que dos trayectorias cercanas se alejan exponencialmente. Se espera que la entropía para un sistema con dinámica caótica sea un número acotado mayor que cero y además invariante.

Sin embargo la estimación de este parámetro no es sencilla, ya que se debe tener un número de observaciones suficientemente grande ($n \rightarrow \infty$) y con una exactitud lo suficientemente grande (lo que dificulta la selección de las particiones $\{R_i\}$ cuando los datos son ruidosos).

4.2 Exponentes de Lyapunov.

En sistemas dinámicos los exponentes de Lyapunov son cantidades que se pueden calcular para definir la forma en que evolucionan las trayectorias de un sistema partiendo de determinadas condiciones iniciales. Por medio de estas cantidades puede determinarse si el sistema convergerá a un atractor simple o a uno extraño.

En general los sistemas dinámicos deterministas poseen igual cantidad de exponentes de Lyapunov como de grados de libertad, y su comportamiento está definido por el mayor de estos exponentes. Cuando el mayor de los exponentes es negativo³ ($\lambda < 0$, es decir, todos son negativos) la trayectoria del sistema dinámico converge a un punto fijo en el espacio de fase; cuando el mayor exponente de Lyapunov es cero ($\lambda = 0$), la trayectoria converge a una órbita periódica⁴ y, por último, para el caso en que el mayor exponente de Lyapunov es mayor que cero ($\lambda > 0$) pueden ocurrir dos cosas. Si $\sum_i \lambda_i > 0$ el sistema es divergente. En el caso en que $\sum_i \lambda_i < 0$ el sistema es acotado pero no se estabiliza ni en una órbita periódica ni en un punto fijo.

Una de las utilidades de los exponentes de Lyapunov es la capacidad de cuantificar la velocidad con que dos trayectorias que parten de una condición inicial muy cercana se acercan o alejan con la evolución del tiempo, es decir, miden el grado de sensibilidad del sistema a variaciones en las condiciones iniciales (Díaz 2000).

Supóngase que se parte de dos condiciones iniciales x_0 y $x_0 + \epsilon_0$ para $t = 0$ y se deja evolucionar el sistema a lo largo del tiempo obteniéndose dos trayectorias. Si se calcula la diferencia entre las trayectorias se tiene que esta evoluciona de manera exponencial de la siguiente forma

$$\epsilon_t \approx \epsilon_0 e^{\lambda t} \quad (4.8)$$

Si el exponente $\lambda < 0$, las trayectorias se combinan en una sola, ya que $e^{\lambda t} \rightarrow 0$ cuando $t \rightarrow \infty$; por el contrario, si el exponente $\lambda > 0$ las trayectorias se alejan exponencialmente una de la otra. Si el sistema es acotado, esta divergencia exponencial es lo que se llama sensibilidad a condiciones

³ En algunos artículos es llamado exponente característico de Lyapunov.

⁴ En caso de presentarse una multiplicidad mayor que uno de este exponente igual a cero la órbita es llamada cuasi-periódica.

iniciales. Naturalmente dos trayectorias no pueden separarse más allá del tamaño del atractor (Kantz y Schreiber 2006).

La cantidad λ es llamada el máximo exponente de Lyapunov ó Exponente característico de Lyapunov y en general es el que determina el comportamiento de las trayectorias del sistema (Strogatz 2000) (Díaz 2000). Los sistemas con dinámica caótica poseen al menos un exponente de Lyapunov positivo, y en sistemas disipativos la suma total de sus exponentes debe ser negativa, de forma que el sistema fluctúa expandiéndose y contrayéndose en una región acotada.

El conjunto de exponentes de Lyapunov es llamado el espectro de Lyapunov, y el conocimiento de este es de gran relevancia debido a que se puede calcular la entropía de Kolmogorov Siani como

$$K = \sum_{\lambda_i > 0} \lambda_i \quad (4.9)$$

En sistemas dinámicos la predictibilidad está restringida por la sensibilidad a condiciones iniciales, esta es una de las razones de importancia para el cálculo del exponente característico de Lyapunov (Sprott 2003). En series de tiempo es posible calcular el mayor exponente de Lyapunov, sin embargo esta tarea ha resultado más compleja de lo que parece.

El algoritmo para la estimación de los exponentes de Lyapunov fue propuesto inicialmente por (Wolf, et al. 1985) sin embargo en (Kantz y Schreiber 2006) y (Small 2005) se advierte sobre tres problemas principales: alta sensibilidad numérica, gran cantidad de datos necesarios para hacer una aproximación a la matriz Jacobiana del sistema y siempre entrega un resultado, incluso cuando sea plica en series aleatorias y el exponente debería ser infinito.

Como lo que se busca es caracterizar la forma en que evolucionan las trayectorias a lo largo del tiempo, existen dos formas de hacer esto, la primera es estimando el exponente de Lyapunov y la segunda es estimando el error de pronóstico no lineal. Aunque estas cantidades son completamente diferentes, cuando se miden a partir de una serie de tiempo su cálculo es muy similar (Small 2005).

El error de pronóstico no lineal en un horizonte de tiempo t para un modelo $f(x, t)$ se define como

$$e_f(t) = \sum_n \|f(x_n, t) - x_{n+t}\| \quad (4.10)$$

donde $f(x_n, t)$ es la evaluación del modelo de pronóstico para la evolución de x_n en el horizonte de tiempo t , y claramente $e_f(t)$ depende de la selección y la clase de modelo f . Luego el error de pronóstico no lineal $E(t)$ es el mínimo modelo tal que

$$E(t) = \min_f e_f(t) \quad (4.11)$$

$E(t)$ debería definirse de una mejor manera, ya que hacer una minimización sobre un conjunto no convexo, no suave y discontinuo de funciones f es impráctico. Por ahora se supondrá que el mejor modelo para el pronóstico x_{n+t} será x_{m_n+t} , donde $\|x_n - x_{m_n}\|$ es mínima, es decir, el pronóstico seguirá la trayectoria del vecino más cercano (esta metodología es similar a la introducida en la sección 3.3, la única diferencia es que en este punto se evalúa el pronóstico dentro de la muestra).

Luego, una nueva definición para el error de pronóstico no lineal es

$$E(t) = \sqrt{\frac{1}{N-t} \sum_{n=1}^{N-t} (x_{n+t} - x_{m_n+t})^2} \quad (4.12)$$

$$m_n = \arg \min_{\substack{1 \leq m_n \leq N-t \\ m_n \neq n}} \|z_n - z_{m_n}\|$$

donde z_n es una representación del embebimiento de x_n . En otras palabras z_{m_n} es el vecino más cercano a z_n y $(x_{n+t} - x_{m_n+t})$ mide su separación en un tiempo t . Suponiendo que el sistema es estacionario y que se cuenta con suficientes datos, este es exactamente el comportamiento que se esperaría del mejor modelo.

Volviendo al cálculo de los exponentes de Lyapunov y siguiendo el método de Wolf, se tiene que se debe evaluar la divergencia de dos orbitas cercanas. En este caso suponemos que solo tenemos una serie de tiempo, por lo tanto al situar un punto de referencia en la serie y se observa hacia el pasado o el futuro, se encontrarán puntos en el espacio cerca del de referencia. Luego se puede aproximar la divergencia de las trayectorias a partir de la divergencia de dos segmentos de la misma serie. Considérese un punto z_n con N_N vecinos cercanos $z_{m_1}, z_{m_2}, \dots, z_{N_N}$. La divergencia de las trayectorias cercanas puede aproximarse como

$$L(z_n, t) = \log \left(\frac{1}{N_N} \sum_{i=1}^{N_N} (x_{n+t} - x_{i+t})^2 \right) \quad (4.13)$$

nótese que la divergencia de las trayectorias se calcula sobre la serie original y no sobre la reconstrucción del espacio de fase. Finalmente se elabora un promedio sobre las divergencias de cada segmento

$$L(t) = \frac{1}{N} \sum_{n=1}^N L(z_n, t) \quad (4.14)$$

De esta forma una región lineal en la gráfica de $L(t)$ es un buen indicador de divergencia exponencial, y su pendiente será el máximo exponente de Lyapunov. Es más si este valor es significativamente diferente (mayor) que cero es un buen indicio de que existe un exponente positivo, sin embargo el cálculo del espectro completo de Lyapunov sigue siendo un inconveniente cuando no se conoce la dinámica del sistema.

4.3 Dimensión de Correlación

Los sistemas disipativos con comportamiento caótico evolucionan temporalmente en atractores extraños, los cuales pueden ser caracterizados con medidas como la dimensión fractal D , la cual es menor que el número de grados de libertad del sistema. Esta medida ha sido ampliamente usada para determinar que tan “extraño” es un atractor (Grassberger y Procaccia 1983) (Eckmann y Ruelle 1985). Sin embargo el cálculo de esta medida ha resultado complejo de implementar, sobre todo cuando $D > 2$ y cuando debe ser calculada a partir de una serie de tiempo finita, considerándose en algunos casos impráctica.

En (Grassberger y Procaccia 1983) se propuso una nueva medida llamada la Dimensión de Correlación, la cual se encuentra estrechamente relacionada con la dimensión fractal, e incluso en algunos casos puede considerarse de mayor relevancia, ya que hace más énfasis en las regiones del espacio de fase más visitadas por las trayectorias, es decir, las regiones con mayor densidad de puntos (Sprott 2003). Se desarrollo como una alternativa al cálculo de la dimensión fractal de forma que fuera posible caracterizar series de datos experimentales. Desde su aparición, este algoritmo ha sido uno de los más utilizado en la literatura.

(Small 2005) define la dimensión de correlación como una extensión de la noción de dimensión a objetos con dimensión fractal. En las dimensiones enteras es fácil establecer que la medida del volumen $V(\varepsilon)$ varía de la siguiente forma:

$$V(\varepsilon) \propto \varepsilon^d \quad (4.15)$$

Donde ε es la longitud en una dimensión, y d es la dimensión del objeto. Para un objeto fractal es común adoptar una relación como la anterior para determinar su dimensión. Luego d estaría dada por

$$d \propto \frac{\log V(\varepsilon)}{\log(\varepsilon)} \quad (4.16)$$

Y de esta forma

$$d \approx \lim_{\varepsilon \rightarrow 0} \frac{\log V(\varepsilon)}{\log(\varepsilon)} \quad (4.17)$$

Supóngase $\{z_n\}_{n=1}^N$ un embebimiento para una serie de tiempo en \mathcal{R}^{d_s} . Se define la función de correlación $C_N(\varepsilon)$ por medio de la función.

$$C_N(\varepsilon) = \binom{N}{2}^{-1} \sum_{0 \leq i < j \leq N} \theta(\|z_i - z_j\| < \varepsilon) \quad (4.18)$$

Donde $\theta(\blacksquare)$ es la función indicadora⁵, la cual tiene un valor de **1** si la condición se cumple ó **0** si no se cumple, $\|\blacksquare\|$ es una función de distancia típica en \mathcal{R}^{d_s} y la suma $\sum_i \theta(\|z_i - z_j\| < \varepsilon)$ es el

⁵ En inglés es llamada *Heaviside Function* ó *Heaviside Step Function*

número de puntos a una distancia menor que ε de z_j . Si los puntos z_i se encuentran distribuidos uniformemente dentro del objeto, esta suma⁶ es proporcional al volumen de la intersección entre una esfera de radio ε con el objeto estudiado, y $C_N(\varepsilon)$ es proporcional al promedio de dichos volúmenes de las esferas. Luego se espera que

$$C_N(\varepsilon) \propto \varepsilon^{d_c} \quad (4.19)$$

Donde d_c es la dimensión del objeto. Luego la dimensión de correlación se define como

$$d_c = \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\log(C_N(\varepsilon))}{\log(\varepsilon)} \quad (4.20)$$

Este planteamiento es una estimación de la probabilidad de que dos puntos tomados aleatoriamente se encuentren dentro de la misma esfera a una distancia menor que ε y se selecciona de esta manera ya que se considera que el concepto de dimensión sigue teniendo sentido en situaciones donde los puntos de muestreo no se encuentren distribuidos uniformemente sobre el objeto.

Dado que más que hacer una aseveración en cuanto a la naturaleza del sistema, la dimensión de correlación da indicaciones de lo que *no* es el sistema, a continuación se presenta la interpretación de esta medida.

El ruido tiene una dimensión de correlación infinita, por lo tanto este tiende a expandirse hasta llenar cualquier espacio de embebimiento. En otras palabras la dimensión de correlación será igual a la dimensión de embebimiento; por otro lado, los sistemas periódicos tienen dimensión entera. Luego, una dimensión de correlación no entera y significativamente menor a la dimensión de embebimiento es un indicador de que el sistema no está dominado por ruido o por una órbita periódica (o superposición de orbitas periódicas).

El método de Grassberger y Procaccia para el cálculo de d_c deduce la pendiente resultante de la grafica de $\log(C_N(\varepsilon))$ contra $\log(\varepsilon)$. Aunque para valores pequeños de ε puede haber gran variabilidad de la pendiente, una grafica típica de la aplicación del método debe mostrar una región en la cual d_c permaneces relativamente constante.

Aunque se mencionó la gran acogida del algoritmo de Grassberger y Procaccia para el cálculo de la dimensión de correlación, (Small 2005) argumenta que se han detectado algunos problemas de ambigüedad en los resultados. Por otro lado según (Kohers, Pandey y Kohers 1997) este algoritmo necesita una gran cantidad de datos para que funcione de manera eficiente y en (Soofi y Cao 1999) se estima que se necesitan al menos 10^n muestras para obtener resultados consistentes con una dimensión de embebimiento de n .

⁶ Llamada la suma de correlación y donde $\lim_{N \rightarrow \infty} C_N(\varepsilon)$ es la integral de correlación (Sprott 2003)

(Judd 1992) asegura que el problema más evidente del algoritmo elaborado por Grassberger y Procaccia es que este cálculo de la dimensión de correlación depende de una elección subjetiva del intervalo en la cual se espera que la pendiente de la gráfica sea estable, ya que una pequeña alteración en la región que debe seleccionarse puede llevar a cambios sustancialmente grandes.

Judd asume que localmente un atractor puede ser modelado como un producto cruz entre un espacio euclidiano y un conjunto de Cantor. Judd demostró que para estos objetos una mejor descripción de $C_N(\varepsilon)$ para $\varepsilon < \varepsilon_0$

$$C_N(\varepsilon) \approx \varepsilon^{d_c} q(\varepsilon) \quad (4.21)$$

Donde $q(\varepsilon)$ es un polinomio de orden t ($t \geq 1$, y en la práctica $t = 1$ es suficiente). A diferencia del método de Grassberger se incluye en $C_N(\varepsilon)$ una dependencia del polinomio $q(\varepsilon)$, el cual puede capturar los cambios en la pendiente. De esta forma se elimina la dependencia de la selección de una región de ajuste por la dependencia del parámetro ε_0 , por lo tanto $d_c = d_c(\varepsilon_0)$ (Small, Yu, y otros 2000).

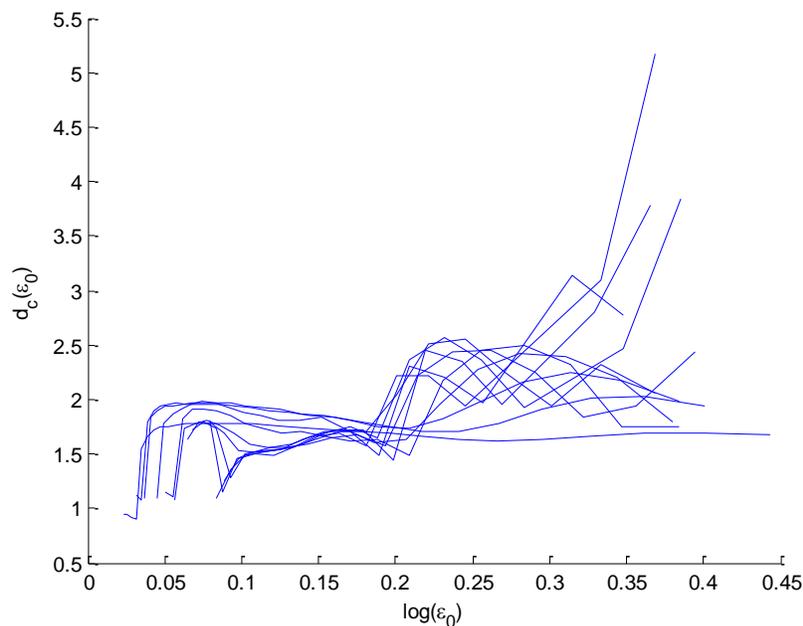


Figura 7: Resultado del cálculo de la dimensión de correlación del sistema dinámico de Lorenz, por medio del algoritmo de Judd.

El resultado de este algoritmo funciona de dos formas diferentes, si se puede observar una región horizontal en esta gráfica esto indica una medida aproximada de la dimensión de correlación como en la figura 7 donde se muestra el resultado del algoritmo de Judd para el sistema de Lorenz, el valor de su dimensión de correlación es de aproximadamente 2.08 (Small 2005).

Por otro lado, este algoritmo presenta un indicador de fractalidad de la serie de tiempo, entendiéndose por fractalidad la autosimilaridad de la serie a diferentes escalas de tiempo y en diferentes dimensiones de embebimiento. Al encontrarse un patrón autosimilar como el encontrado en la figura 8, (donde se muestra el resultado del algoritmo de Judd aplicado al sistema dinámico de Chua), esto indica un comportamiento fractal en la serie, típico de procesos con dinámica caótica embebida.

Existe otro algoritmo adicional para el cálculo de la dimensión de correlación explicado en (Yu, y otros 2000) y (Small 2005) llamado Algoritmo de Kernel Gaussiano (GKA)⁷. Este propone una modificación en el cálculo de la dimensión de correlación introducido por Grassberger y Procaccia reemplazando la función indicadora $\Theta(\|z_i - z_j\| < \varepsilon)$ por una función continua la cual, como el nombre del método lo indica, puede ser la función kernel Gaussiana $\exp\left(-\frac{\|z_i - z_j\|^2}{4\varepsilon}\right)$. Luego, haciendo una generalización de la integral de correlación se tiene la ecuación (4.22)

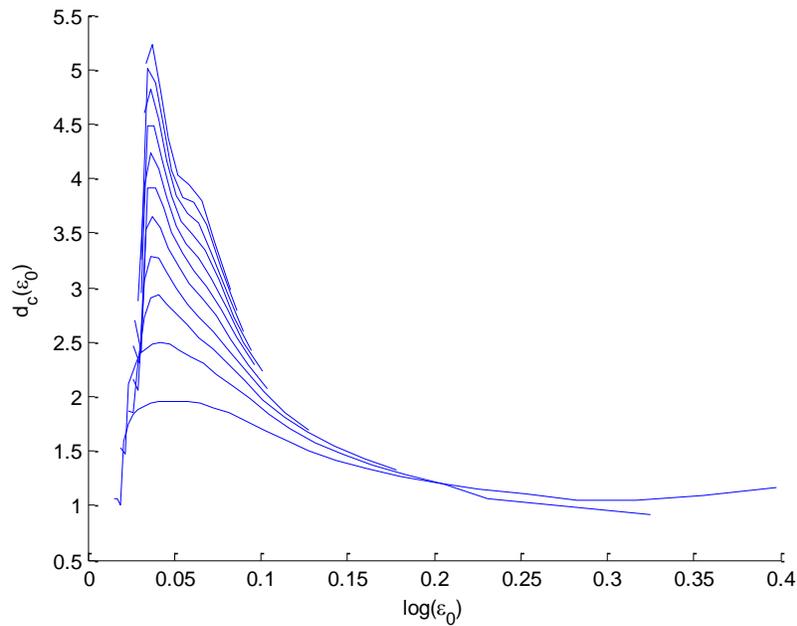


Figura 8: Resultado del cálculo de la dimensión de correlación del sistema dinámico de Chua, por medio del algoritmo de Judd.

$$T_N^{(d_c)}(h) = \frac{1}{N} \sum_{0 \leq i \leq N} \left(\frac{1}{N-1} \sum_{\substack{0 \leq j \leq N \\ j \neq i}} \exp\left(-\frac{\|z_i - z_j\|^2}{4h^2}\right) \right) \quad (4.22)$$

⁷ Kernel Gaussian Algorithm es el nombre en inglés de este método.

En este caso h cumple la misma función que ϵ en el algoritmo de Judd. Para cualquier función kernel se puede demostrar que la ley de escalamiento de esta función de correlación puede describirse como

$$T_N^{(d_g)}(h) \propto h^{d_g} \quad (4.23)$$

Sin en lugar de la función de kernel Gaussiana se escoge $\phi(h) = \exp\left(-\frac{x^2}{4}\right)$, se tiene que

$$T_N^{(d_g)}(h) \approx \phi\left(\frac{h^2}{h^2 + \sigma^2}\right)^{d_g} e^{(-K\tau_g d_g) \left(\frac{h^2 + \sigma^2}{d_g}\right)^{\frac{d_g}{2}}} \quad (4.24)$$

donde τ_g y d_g son los parámetros de embebimiento, K es la entropía de correlación y σ es la desviación estándar del ruido medido en la serie de tiempo. La ventaja de este método radica en que permite calcular simultáneamente la K , σ y lo más importante, la dimensión de correlación d_g . Además el método es robusto al ruido si este se considera aditivo y gaussiano. En este caso un área horizontal ó “plana” en la grafica del algoritmo GKA provee un valor aproximado de la dimensión de correlación. Además como se estableció anteriormente la entropía debe converger a un valor positivo.

En este capítulo se mostraron las medidas invariantes más ampliamente utilizadas para la determinación de la existencia de caos determinista en series de tiempo. Estas medidas será utilizadas en el análisis de la series de la tasa representativa del mercado (TRM) de Colombia que se desarrollará más adelante. En el siguiente capítulo se mostrarán algunas medidas estadísticas que se utilizan para descartar la naturaleza estocástica de las series de tiempo, de esta forma el análisis se considerará completo al poder determinar si las series son o no estocásticas en su totalidad.

5. Análisis estadístico por medio del método *Surrogate Data*

Como se afirmó en el capítulo anterior, aunque se han desarrollado diversos métodos para el cálculo de los diferentes invariantes, se sabe que estas estimaciones no son cien por ciento confiables. Con la presentación de cada método se han destacado los posibles problemas que puede presentar cada algoritmo en presencia de ruido excesivo o incluso una cantidad de datos insuficiente para capturar las dinámicas no lineales en caso de existir en la serie.

Los procesos estocásticos lineales pueden generar series de datos que a simple vista parecen tener gran complejidad en su comportamiento, sin embargo no todas las estructuras que se hallan en las series de tiempo son derivadas de una naturaleza no lineal en el sistema. Más aún, es posible que el cálculo de invariantes, como en el caso del máximo exponente de Lyapunov, genere un resultado que supone la presencia de caos determinista cuando las dinámicas son generadas en su totalidad por un sistema lineal. Se hace necesario entonces adicionar una medida de certidumbre a las conclusiones que se deriven de los resultados obtenidos en las herramientas de análisis.

Supóngase que se calcula una medida no-lineal ψ de los datos, por ejemplo la dimensión de correlación, el máximo exponente de Lyapunov, el error de pronóstico no lineal, o cualquier otro. Para saber si el valor de ψ sugiere que los datos son producidos por dinámicas no lineales, primero se debe saber que distribución puede obtenerse de los valores de ψ para un modelo lineal comparable. Si el valor calculado sobre la serie no es consistente con la descripción lineal, los datos podrían ser no lineales (Kantz y Schreiber 2006).

El método de análisis de linealidad por medio de Surrogate data fue propuesto por James Theiler en su artículo (Theiler, Eubank, y otros 1992), y permite identificar si las dinámicas son consistentes con ruido filtrado linealmente ó con un sistema dinámico no lineal. Este se basa en el cálculo de una prueba de hipótesis mediante la definición de una hipótesis nula, la generación de conjuntos de Surrogate data⁸ que son consistentes con la hipótesis nula y finalmente el cálculo de un estadístico discriminante para la serie original de datos y los conjuntos de Surrogate data. (Theiler, Eubank, y otros 1992) describe este método como una aplicación del método de Bootstrapping para la generación de series de tiempo a partir de un conjunto de datos y una distribución supuesta.

Como es común, debe iniciarse desde el grado más bajo de complejidad hasta el más alto. Luego, los tres tipos de hipótesis nulas que se evaluarán son: 1) los datos son resultado de un proceso independiente e idénticamente distribuido, 2) los datos son resultado de un proceso aleatorio sometido a un filtrado lineal (por ejemplo alguna variación del proceso ARMA), y por último 3) los

⁸ El autor no encontró ninguna traducción comúnmente usada en español para este término. Incluso se encontró un artículo en español que llama los conjuntos de esta misma forma (Carbajal, y otros 2001).

datos son resultado de una transformación no lineal de un proceso aleatorio sometido a un filtrado lineal.

Cada una de estas hipótesis necesita generar un conjunto de Surrogate data, que debe realizarse de forma diferente, a continuación se enumeran los métodos de realización que concuerdan respectivamente con la enumeración anterior: 1) mezclar aleatoriamente los datos de la serie original, 2) mezclar aleatoriamente las fases de la transformada de Fourier de la serie de tiempo original, 3) aplicar el mezclado de fases a un proceso aleatorio Gaussiano de amplitud ajustada.

Existe la posibilidad de que los datos a analizar presenten una fuerte tendencia periódica, por lo cual el cálculo de los Surrogates debe hacerse de forma un poco diferente. En lugar de mezclar los datos como en el caso número uno, se divide la serie en los ciclos y estos bloques de datos se mezclan aleatoriamente.

Los algoritmos para el cálculo de estos conjuntos de series fueron introducidos por Theiler en sus artículos (Theiler, Eubank, y otros 1992) y (Theiler 1995), además son ampliamente explicados en (Small 2005), por lo tanto no se profundizará más en este tema. Lo que interesa en este punto es la construcción de las pruebas estadísticas de forma que se pueda concluir a partir de los resultados que estas generan.

5.1 Pruebas estadísticas

Sea ϕ una hipótesis específica, y F_ϕ el conjunto de todos los procesos consistentes con esa hipótesis. Sea $Z \in R^N$ una serie de tiempo y $T: R^N \rightarrow U$ un estadístico que se utilizará para probar la hipótesis ϕ que dice que Z fue generada por un proceso $F \in F_\phi$. Los surrogate data S son generados a partir de la serie Z de la misma longitud y consistentes con la hipótesis ϕ . Se tiene que $U \subseteq R$ y se puede discriminar entre la serie Z y los surrogate data S_i consistentes con ϕ según la densidad de probabilidad estimada $p_{T,F}(t) = P(T(S_i) < t)$, es decir, la densidad de probabilidad de T dado F .

Luego, el proceso para la evaluación estadística puede describirse en unos breves pasos: Se plantea la hipótesis nula como cualquiera de las anteriormente mencionadas; se elaboran los conjuntos de surrogate data a partir de la serie original con los métodos propuestos según la hipótesis nula y utilizando el método de simulación Montecarlo; se calcula el estadístico discriminante para el conjunto de surrogate data y se le ajusta una distribución; se calcula el valor del estadístico discriminante para los datos originales y se evalúa la significancia de este valor con la distribución ajustada.

Si Q_D es el valor del estadístico para la serie de tiempo original, Q_{H_i} es el estadístico para la i -ésima serie obtenida por medio de surrogate data bajo la hipótesis nula, y considérese μ_H y σ_H respectivamente la media y desviación estándar (muestrales) de la distribución de Q_H , entonces (Theiler, Eubank, y otros 1992) define una medida de significancia S como

$$S = \left| \frac{Q_D - \mu_H}{\sigma_H} \right| \quad (5.1)$$

Según Theiler comúnmente Q_H sigue una distribución gaussiana, por lo que el valor p de la prueba puede calcularse como $p = \text{erf}(S/\sqrt{2})$, que es la probabilidad de observar un valor de significancia S o mayor si la hipótesis nula es cierta. (Theiler, Eubank, y otros 1992) asegura que resultado numéricos han demostrado que en la mayoría de los casos Q_H sigue una distribución gaussiana, sin embargo esto no siempre es cierto.

Debido a que no siempre Q_H se distribuye de manera gaussiana, (Small 2005) aconseja utilizar distribuciones adicionales, es por esto que se propone el uso de pruebas estadísticas como por ejemplo la comparación directa entre las medidas seleccionadas ó por medio de la selección de una prueba estadística de bondad de ajuste, como la prueba de Kolmogorov – Smirnov ó la prueba χ^2 de Pearson.

En cuanto al valor estadístico discriminante existen muchas medidas que podría aplicarse con este fin. Debido a que se quiere identificar comportamiento no lineal ó caos en la serie de tiempo las medidas más lógicas a proponer pueden ser la dimensión de correlación, el máximo exponente de Lyapunov ó el error de pronóstico no lineal. Se debe tener en cuenta que la selección de esta medida debe hacerse con cuidado, ya que la efectividad y la potencia de discriminación de la prueba de hipótesis depende de la confiabilidad de estimación de la medida. En (Small 2005) se recomienda seleccionar una medida relacionada con la dimensión de correlación, debido a que es una medida de gran relevancia y ha sido ampliamente usada.

5.1.1 Prueba Kolmogorov – Smirnov

Es una prueba estadística no paramétrica que puede ser utilizada para comparar distribuciones. Si se selecciona como medida discriminante la función de correlación⁹ $C_N(\varepsilon)$ también llamada distribución de distancias entre puntos, la cual fue introducida en la sección 4.3, la prueba puede ser calculada hallando la mayor distancia entre las distribuciones de la siguiente manera

$$\max_{\varepsilon} |C_N(\varepsilon) - \hat{C}_N(\varepsilon)| \quad (5.2)$$

Donde $C_N(\varepsilon)$ es la distribución de distancias entre puntos de la serie original y $\hat{C}_N(\varepsilon)$ es la distribución de distancias entre puntos de los datos de cada conjunto de surrogate data, lo cuales se supone que concuerdan con la hipótesis nula.

La hipótesis nula de la prueba es $H_0 =$ "Las muestras provienen de la misma distribución" (en el caso de comparar dos distribuciones), ó $H_0 =$ "Las muestras provienen de la distribución de

⁹ Como se describió en la sección 4.3, es la probabilidad de que dos puntos z_i y z_j se encuentren a una distancia menor de ε entre sí.

referencia" (en el caso de una sola muestra). Esta prueba se encuentra programada en Matlab y entrega un resultado igual a cero cuando se acepta la hipótesis nula y de 1 cuando se rechaza.

5.1.2 Prueba χ^2

La prueba χ^2 es considerada una prueba no paramétrica y puede usarse para medir la diferencia entre una distribución observada $\hat{C}_N(\varepsilon)$ y una distribución propuesta $C_N(\varepsilon)$ de distancias entre puntos. El test χ^2 asume una distribución discreta propuesta con su respectiva división en intervalos y la compara con la distribución de un conjunto de observaciones experimentales mediante el siguiente cálculo:

$$\chi^2 = \sum_i \frac{(N_i - np_i)^2}{np_i} \quad (5.3)$$

Donde p_i denota la probabilidad esperada que tiene una distancia entre puntos aleatoria de caer en el i -ésimo intervalo de la distribución propuesta $C_N(\varepsilon)$, N_i es el número de distancias entre puntos pertenecientes al i -ésimo intervalo de la distribución observada $\hat{C}_N(\varepsilon)$ y n es el número de distancias entre puntos.

En este caso la hipótesis nula es $H_0 =$ "No hay diferencia significativa entre las frecuencias propuestas y las frecuencias observadas". El cálculo de esta prueba se desarrollará mediante un algoritmo tomado de "Numerical Recipes in C" y programado por el profesor Ph D. Michael Small de la Universidad Politécnica de Hong Kong. El cual devuelve el valor del estadístico χ^2 y la probabilidad de obtener este valor del estadístico dada la hipótesis nula.

En esta sección se mostraron algunas pruebas estadísticas adicionales que pueden aportar información relevante a las posibles conclusiones del cálculo de invariantes. En la siguiente sección se mostrará la aplicación de las herramientas, revisadas en este trabajo, a un caso real como es el análisis de la serie de tiempo de la tasa representativa del mercado colombiano (TRM).

6. Análisis de datos

El propósito principal de este trabajo a parte de la recopilación de las técnicas utilizadas para el análisis no lineal de la información, es el análisis de una de las series de tiempo más relevantes para la economía colombiana como el precio de la tasa representativa del mercado (TRM) a escala diaria desde el 2 de diciembre de 1991 hasta 21 de diciembre de 2009.

A continuación se describirá el procedimiento que se llevará a cabo para el análisis. Antes de comenzar con cualquier análisis se calculan los rendimientos logarítmicos de los precios y se someten a la prueba de normalidad de Kolmogorov – Smirnov para descartar que se traten de datos aleatorios normalmente distribuidos.

La primera prueba a desarrollar será el análisis de reescalamiento hallando el exponente de Hurst de la serie junto con su valor esperado y el estadístico V. Para desarrollar este análisis se debe aplicar un proceso de prewhitening sobre la serie por medio de un filtro $AR(1)$ de forma que se elimine parte del sesgo causado por las dependencias lineales (Peters 1994). Este análisis se desarrolla con el fin de detectar persistencia o anti-persistencia en la serie y determinar si existen ciclos pronunciados en la serie.

Se realizará el análisis de embebimiento de forma que puedan determinarse los parámetros τ_{ϵ} y d_{ϵ} del mejor embebimiento posible para la serie. De una buena reconstrucción del espacio de fase depende en gran parte el desempeño de los algoritmos de análisis posteriores y el pronóstico de la variable. Adicionalmente se mostrará el pronóstico desarrollado a partir de la dinámica reconstruida por medio de un modelo constante local (LCM) y sus resultados se compararán con los de un modelo ARIMA.

Finalmente elaborará el cálculo de invariantes de la serie de tiempo y el análisis por medio de la metodología de surrogate data de forma que puedan sacarse conclusiones sobre la serie de la TRM en cuanto a la presencia de caos determinista o comportamiento no lineal de las series.

6.1 Análisis de la tasa representativa del mercado (TRM)

6.1.1 Los datos.

A continuación se presenta el análisis de la tasa representativa del mercado tomada desde el 2 de diciembre de 1991 hasta el 19 de diciembre de 2009. Se eliminaron los datos replicados en los fines de semana de forma que solo se consideran los días bursátiles. Esta serie se dividió en dos partes, la primera será utilizada para el desarrollo de todas las pruebas de caracterización y constará de 4000 muestras, la segunda parte será utilizada para la validación del modelo de pronóstico y consta de 341 muestras. A continuación en la figura 9 se muestra la grafica de la serie de tiempo de la TRM de los datos para el análisis.

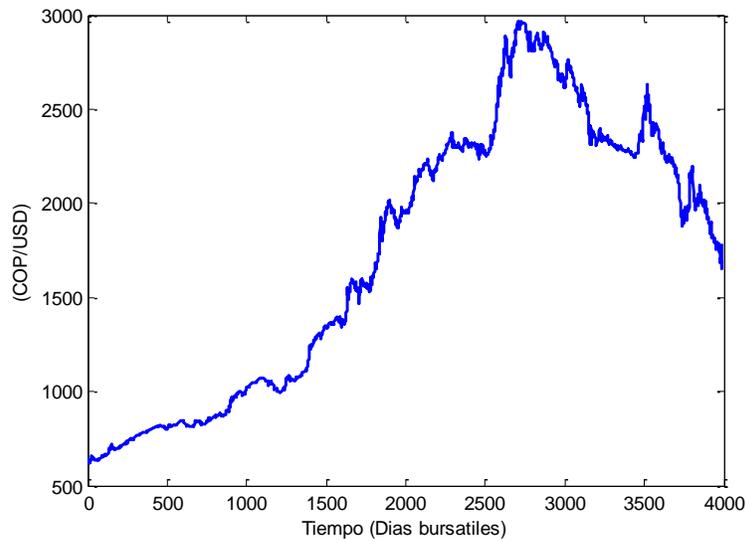


Figura 9: Serie de tiempo de la TRM en días bursátiles.

Al calcular los rendimientos logarítmicos se obtiene la figura 10. Claramente se muestra que los rendimientos logarítmicos presenta el fenómeno de heterocedasticidad. En adelante el análisis se desarrollará sobre la serie de rendimientos logarítmicos.

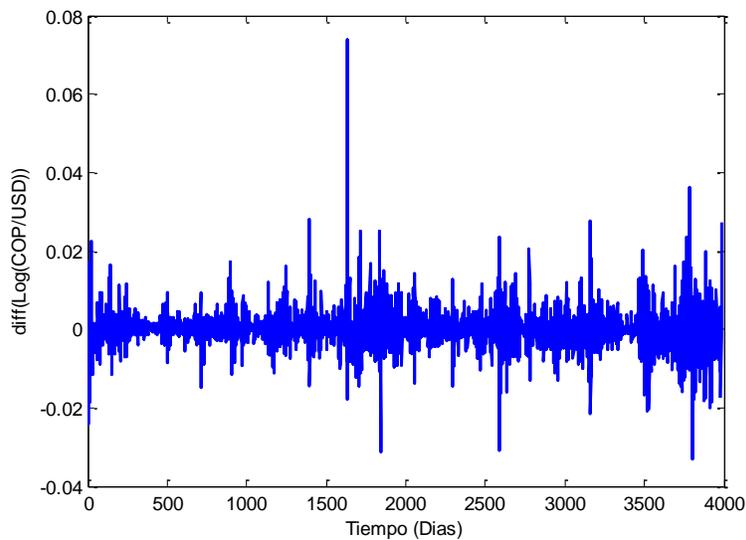


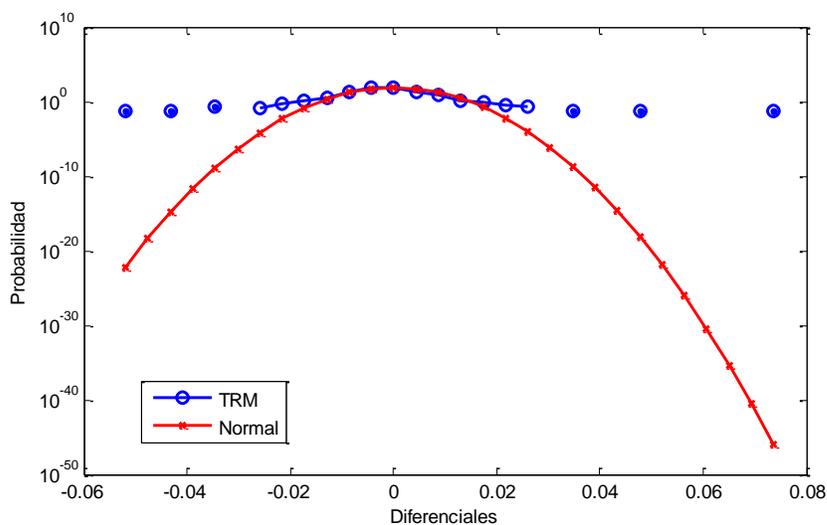
Figura 10: Serie de tiempo de los rendimientos logarítmicos de la TRM en días bursátiles.

La tabla muestra los estadísticos básicos calculados sobre la serie de rendimientos logarítmicos de la TRM, los cuales muestran una desviación estándar pequeña y una curtosis muy elevada, lo cual indica una aglomeración de los datos alrededor de la media. Por otro lado se muestra un valor de asimetría positivo, lo que indica que la distribución de los datos cuenta con una cola más larga hacia la derecha.

Estadístico	Valor
Media	2.507 e-4
Desviación estándar	0.0049
Asimetría	0.8068
Curtosis	25.9272

Tabla 1: Estadísticos básicos para los rendimientos logarítmicos de la TRM

La Figura 11 compara el histograma de frecuencias de los rendimientos logarítmicos de la TRM con la distribución normal. La diferencia entre las dos funciones de distribución de probabilidad (FDP) es mayor en las colas que en valores cercanos a la media, es por esto que el eje vertical (la Probabilidad) se muestra en escala Logarítmica en vez de escala lineal. Se puede ver como estos rendimientos presenta un fenómeno de colas pesadas. Lo que hace que una suposición de normalidad sobre los datos no sea muy



acertada.

Figura 11: Comparación del histograma de rendimientos logarítmicos de la TRM con la distribución normal.

Para corroborar los resultados de la Figura 11, se decide realizar la prueba de normalidad de Kolmogorov – Smirnov en la cual se plantea como hipótesis nula:

$$H_0: \text{Los datos siguen una distribución normal}$$

Dando como resultado el rechazo de la hipótesis nula. Del mismo modo se evalúa la prueba de Jarque – Bera para el cual se tiene la siguiente hipótesis nula:

$$H_0: \text{Los datos siguen una distribución normal con desviación típica y media desconocidas}$$

En este caso también se obtiene como resultado el rechazo de la Hipótesis nula.

6.1.2 Análisis de reescalamiento (R/S)

Para este análisis se aplica el proceso de prewhitening a la serie de retornos logarítmicos de la TRM por medio de un filtro $AR(1)$ de forma que se elimine parte del sesgo causado por las dependencias lineales (Peters 1994). El cálculo del exponente de Hurst resulta de la pendiente de la gráfica que se muestra en la figura 11 y su ajuste lineal produce el siguiente resultado: $f(x) = 0.6188 * x - 0.1459$ y por lo tanto $H = 0.6188$. El ajuste del valor esperado $E(R/S)$ es: $f(x) = 0.5115 * x + 0.05079$, luego $E(R/S) = 0.5115$.

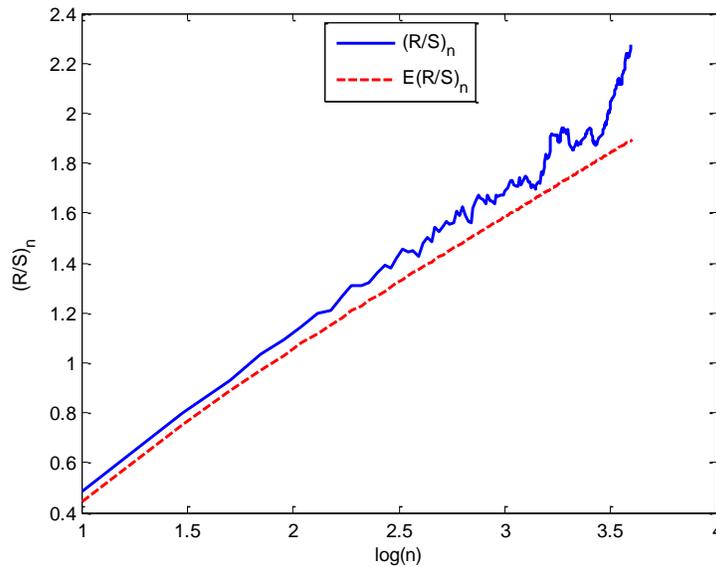


Figura 12: Cálculo del exponente de Hurst. Claramente se observa como los rendimientos logarítmicos de la TRM presentan un crecimiento más pronunciado que un proceso aleatorio.

Para hallar que tan significativa es la diferencia del exponente calculado con respecto al valor esperado se debe calcular cuantas desviaciones estándar constituyen esta diferencia. Calculando la varianza se tiene que.

$$Var(H) = \frac{1}{T} = \frac{1}{4000} = 0.00025 \quad (6.1)$$

Luego, la desviación estándar es $\sigma = \sqrt{Var} = 0.0158$. Haciendo la comparación se tiene que:

$$H = 0,6188 > 0.5425 = E(R/S) + 2.58 * \sigma \quad (6.2)$$

Por lo tanto se puede decir que la serie presenta un comportamiento persistente con una confianza del 99%, lo cual indica que la serie de rendimientos logarítmicos presenta un fenómeno de memoria larga.

Continuando con el análisis se procede a calcular el estadístico V a partir del cálculo del análisis de reescalamiento. En el resultado presentado en la figura 12, el primer cambio de pendiente significativo aparece en $n = 130$, el cual podría utilizarse para el cálculo del rezago óptimo de embebimiento. En cuanto a los demás cambios significativos de pendiente de la grafica se encuentran en rezagos del orden de los 1300 datos en adelante, lo cual puede reflejar variaciones debido a ciclos económicos que afectan la tasa de cambio a largo plazo.

Hasta este punto se tiene que la serie de los retornos logarítmicos de la TRM presentan un comportamiento persistente, lo que indica que recorre un rango más amplio que un movimiento browniano geométrico, o en otras palabras presenta un comportamiento típico de series de tiempo con memoria larga. Además el estadístico V refleja cambios de pendiente concentrados en el largo plazo lo que puede evidenciar influencia de ciclos económicos en la serie de la TRM.

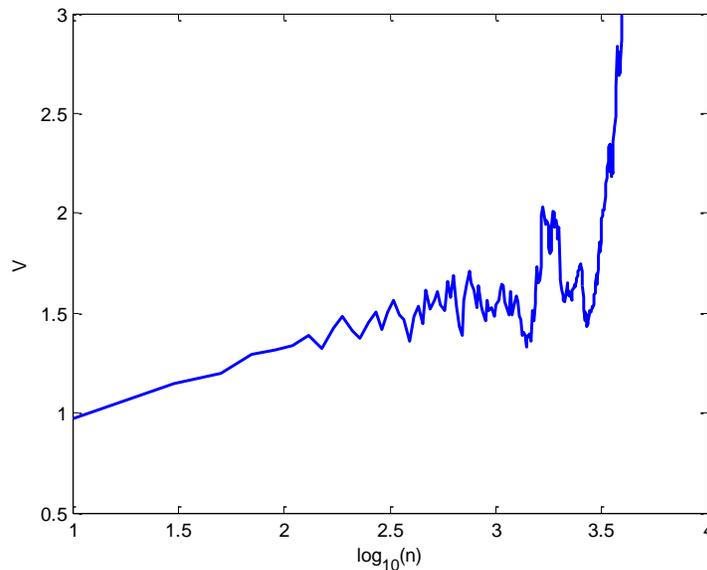


Figura 13: Estadístico V para la TRM.

6.1.3 Teoría de embebimiento

A continuación se procede con la reconstrucción de las dinámicas inmersas en la serie de tiempo. Como se mostró en capítulos anteriores, para este propósito se deben calcular dos parámetros específicos, el primero es el rezago óptimo de embebimiento τ_e , y el segundo es la dimensión de embebimiento d_e .

Para el cálculo del rezago óptimo de embebimiento τ_e se cuenta con varias opciones: puede seleccionarse como el primer valor cercano a cero en el cálculo de la función de autocorrelación, puede seleccionarse como el primer mínimo local de la función de información mutua ó puede calcularse como una cuarta parte del periodo fundamental de la serie. En este caso, calculando la función de autocorrelación (figura 13) se tiene que su primer valor más cercano a cero es en $\tau = 4$, y calculando la función de información mutua (figura 14) se tiene que su primer mínimo

local es en $\tau = 4$. Como en este caso no se tiene certeza de cuál es el periodo fundamental de la serie y se cuenta con un resultado redundante en las dos medidas descritas, se selecciona el rezago óptimo como $\tau_e = 4$.

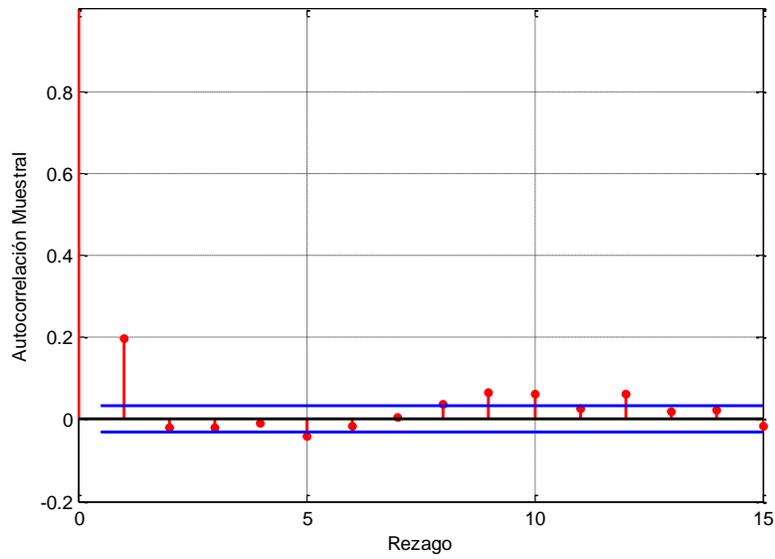


Figura 14: Función de autocorrelación para la serie de incrementos logarítmicos de la TRM.

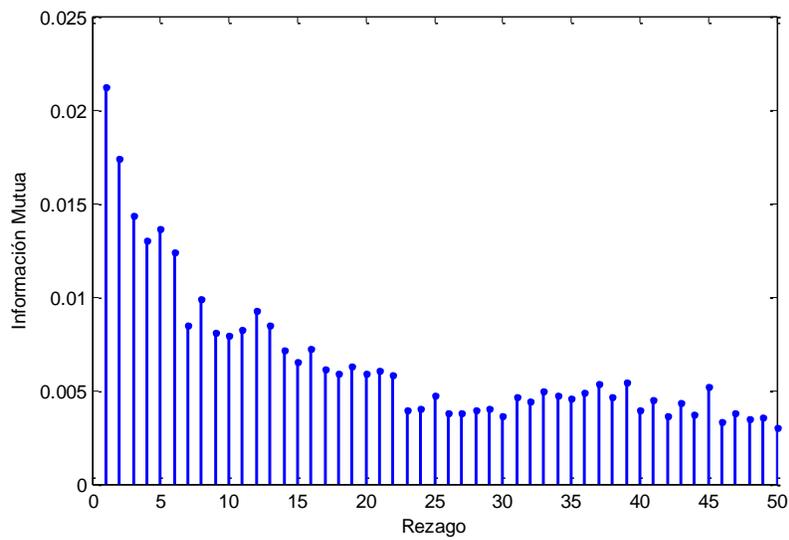


Figura 15: Función de información mutua promedio para la serie de incrementos logarítmicos de la TRM

Ahora, para el cálculo de la dimensión de embebimiento d_e se utilizará el algoritmo propuesto por Cao, el cual utiliza el valor de τ_e y el método de los falsos vecinos cercanos. La figura 15 muestra la convergencia del método al valor $d_e = 7$.

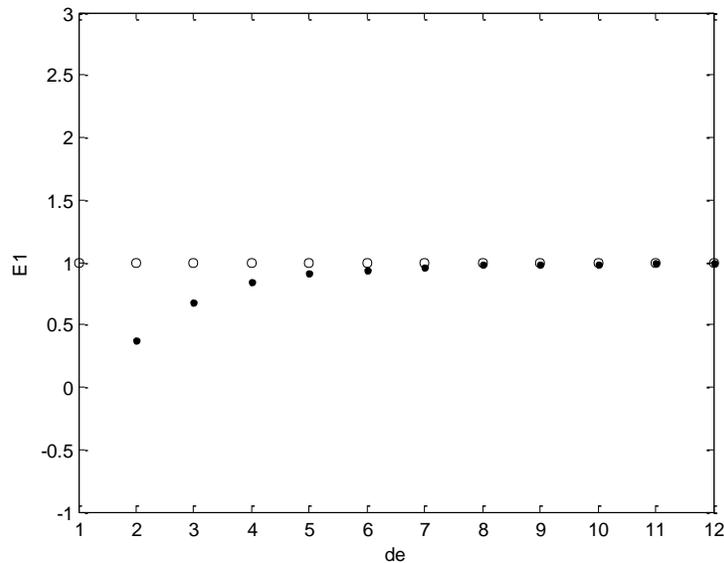


Figura 16: Convergencia de la dimensión de embebimiento d_e calculado por medio del método de Cao.

Se debe tener en cuenta que es posible seleccionar una dimensión de embebimiento mayor, sin embargo esto trae dos desventajas: 1) puede ocasionar un incremento computacional en las herramientas de análisis que requieran del embebimiento; 2) puede incrementar de manera innecesaria la complejidad del embebimiento, tratando de modelar dinámicas que puede ser consideradas como parte ruido (también llamado sobre – ajuste), es decir aumentando la cantidad de ruido presente en el sistema. Mientras más sencillo pueda hacerse el embebimiento mejor (Kantz y Schreiber 2006).

Al tener los parámetros del modelo de reconstrucción del espacio de fase estimados se procede a elaborar dos modelos de pronóstico para los retornos logarítmicos de la TRM, por medio de un Modelo Local de Predicción Constante (LCM), el cual se describió en la sección 3.3 y un modelo de redes neuronales con estructura de perceptrón multicapa el cual será descrito a continuación.

Se ajustó un modelo de red neuronal en forma de perceptrón multicapa con 20 neuronas en la única capa oculta. La ecuación (6.3) determina este modelo.

$$y(k) = \sum_{j=1}^m w_j \cdot f_j(\sum_{i=1}^n a_{ij} \cdot x_i(k) + b_j) + c \quad (6.3)$$

donde $y(k)$ equivale a la variable de salida, el cual se eligió como el pronóstico del retorno logarítmico de la TRM; $x_i(k)$ son las entradas al modelo en el tiempo k , que en este caso equivalen a la matriz de rezagos construida a partir de d_e y τ_e ; y finalmente w_j , a_{ij} , b_j y c son los parámetros del modelo, los cuales se ajustaron por medio del método de regularización bayesiana del toolbox de redes neuronales de Matlab. La figura 17 muestra la estructura del modelo.

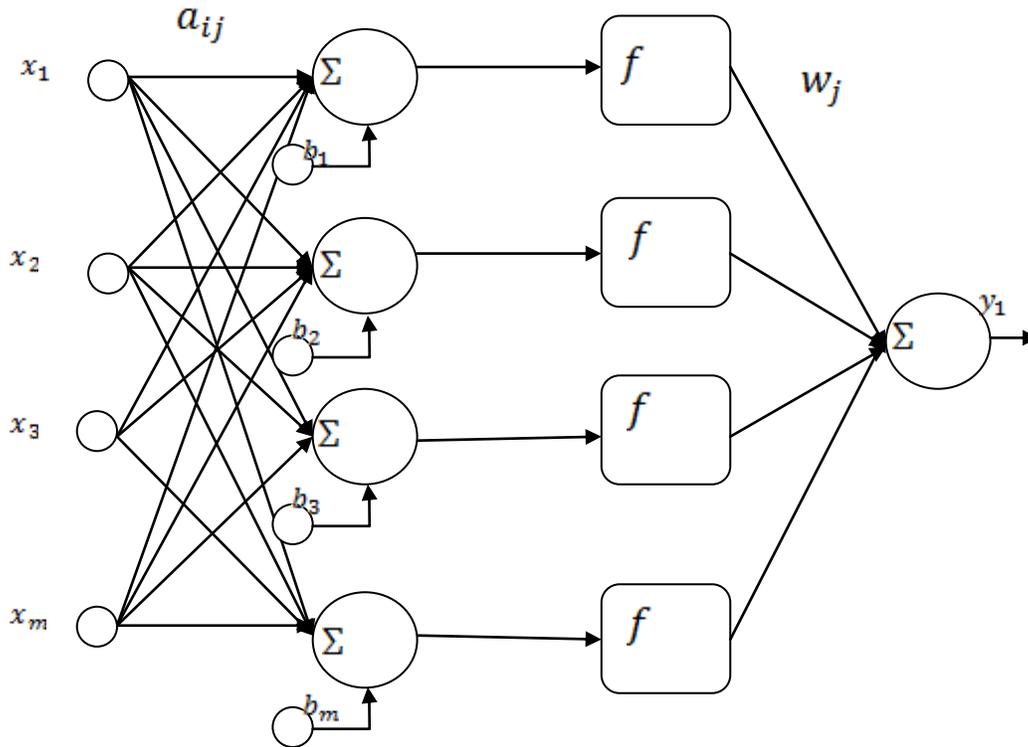


Figura 17: Estructura esquemática del perceptrón multicapa.

Los resultados del modelo LCM fueron contrastados con los resultados de un modelo ARIMA ajustado con el paquete de software R mediante la función autoarima. El resultado del ajuste con sus parámetros se muestra en la tabla 2, donde los parámetros obtenidos son de un modelo $ARIMA(1,1,3)$, cuya estructura está dada por la ecuación (6.3):

$$p[i] = ar1 * p[i - 1] + ma1 * a[i - 1] + ma2 * a[i - 2] + ma3 * a[i - 3] \quad (6.3)$$

Parámetro	Valor	SE
ar1	0.8519	0.0286
ma1	-1.6673	0.0338
ma2	0.5013	0.0407
ma3	0.1664	0.0156

Tabla 2: Resultado del ajuste de un modelo **ARIMA(1,1,3)** para los rendimientos logarítmicos de la serie de la TRM por medio del paquete estadístico R.

Se utilizaron la medida RMSE para la estimación de la bondad del pronóstico en los retornos logarítmicos y los estimadores MAPE y RMSE para la estimación de la desviación equivalente en la serie de los precios. El ajuste proporcionado por los modelos descritos anteriormente sobre los retornos logarítmicos se muestra en la tabla 3, y el ajuste resultante en la serie de tiempo del precio de la TRM se muestra en la tabla 4.

	LCM	ARIMA	Red Neuronal
RMSE	0.0142	0.0121	0.0121

Tabla 3: Comparación de resultados pronostico de los residuales de la TRM con los modelos LCM, ARIMA y una Red Neuronal

	LCM	ARIMA	Red Neuronal
RMSE	31.19	26.47	26.51
MAPE	1.05%	0.90%	0.91%

Tabla 4: Comparación de resultados pronostico de los precios de la TRM con los modelos LCM, ARIMA y una Red Neuronal aplicados en los residuales

Si se tiene en cuenta que la variación diaria promedio de la serie de la TRM es de 20 pesos, y el mejor de los pronósticos presenta un error **RMSE = 26.47** pesos, se puede concluir que en realidad ninguno de los modelos es suficiente para generar un pronóstico aceptable a un paso. Claramente el modelo LCM es el que menor desempeño tiene en la captura de la dinámica que quiere extraerse de los residuales, mientras que el modelo ARIMA(1,1,3) y el modelo de redes neuronales obtuvieron desempeños similares. Este resultado sugiere que el embebimiento hallado para la serie de tiempo no captura correctamente la dinámica subyacente de la serie ó no fue correctamente explotado, por lo cual no se producen mejores resultados que una herramienta de modelamiento lineal como es el modelo ARIMA.

Con los resultados obtenidos hasta el momento no se pueden sacar muchas conclusiones, solo se tiene que la serie presenta un comportamiento de persistencia que indica la presencia de un fenómeno de memoria larga, pero este no fue satisfactoriamente capturado y aprovechado para la elaboración de un pronóstico por medio de los modelos seleccionados. Debido a esto se propone a continuación el cálculo de algunas cantidades invariantes, las cuales pueden contribuir con pistas adicionales para la explicación del comportamiento de la serie.

6.1.4 Cálculo de invariantes

La caracterización del sistema es de suma importancia para saber el tipo de dinámica es la que está intentando modelarse, es por esto que el siguiente paso en el análisis es la estimación de invariantes los cuales, según sus resultados, pueden sugerir la presencia de dinámica no lineal, comportamiento caótico e incluso alta presencia de ruido en la trayectoria.

El primer invariante a analizar es máximo exponente de Lyapunov o exponente característico de Lyapunov. Este fue aproximado mediante la implementación de las ecuaciones (4.13) y (4.14) para

una ventana de tiempo variable desde 1 hasta 20 y una dimensión de embebimiento desde 3 hasta 12. La Figura 18 muestra las líneas correspondientes a cada una de las dimensiones de embebimiento, siendo la línea inferior la correspondiente a la dimensión $d_e = 3$ y la línea superior la correspondiente a $d_e = 12$. Como se señaló en la sección 4, la pendiente de la gráfica de $L(t)$ puede ser una aproximación para el máximo exponente de Lyapunov, es por esto que se ajusta un polinomio lineal a cada una de las curvas de $L(t)$ de forma que pueda determinarse una estimación de la pendiente. La estimación de la pendiente de cada línea junto con los límites de confianza al 95% se muestran en la tabla 5.

d_e	Pendiente	Límite Inferior (95%)	Límite Superior (95%)	R^2
3	0.007681	0.006345	0.009016	0.8903
4	0.008201	0.006546	0.009856	0.8576
5	0.008606	0.006833	0.01038	0.8524
6	0.009646	0.007803	0.01149	0.8704
7	0.009223	0.007723	0.01072	0.9027
8	0.009871	0.008531	0.01121	0.9301
9	0.01025	0.008909	0.01159	0.9346
10	0.01107	0.009764	0.01237	0.9464
11	0.01125	0.009855	0.01264	0.9413
12	0.01139	0.01001	0.01277	0.9437

Tabla 5: Ajustes de las pendientes de $L(t)$ para cada d_e desde 3 hasta 12

La tabla muestra el ajuste de los polinomios lineales para cada una de las curvas $L(t)$.

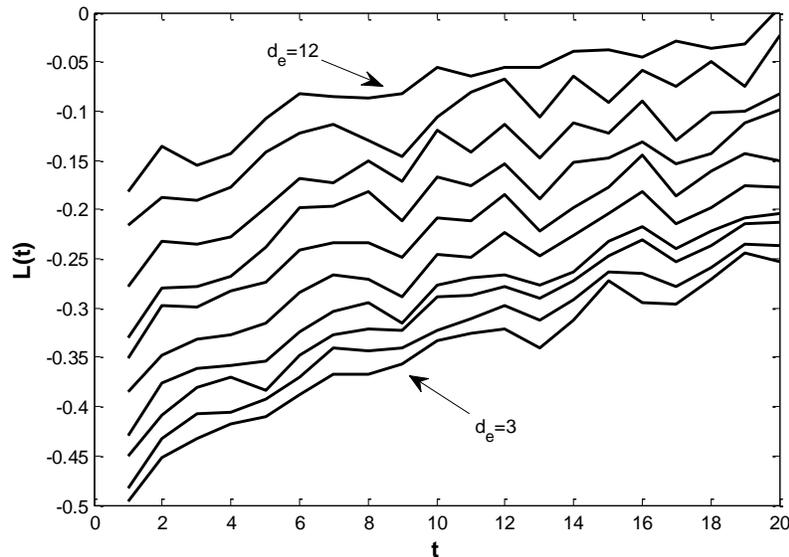


Figura 18: Graficas de $L(t)$ para la estimación del máximo exponente de Lyapunov.

De la figura 18 y la tabla 5 se puede decir varias cosas. Primero la variación de la pendiente del ajuste calculado para cada curva de datos $L(t)$ es muy pequeña, además inicialmente se tiene que todas las pendientes son positivas, luego se podría pensar en un resultado satisfactorio en cuanto a la estabilidad del invariante frente al cambio en la dimensión de embebimiento.

Sin embargo, por otro lado se debe tener en cuenta que la máxima pendiente calculada de la en los ajustes lineales es alrededor de 0.011 , lo cual no es significativamente diferente a cero. Por lo cual, a pesar de mostrarse un valor de divergencia exponencial de las trayectorias no es conveniente concluir positivamente en cuanto a la presencia de caos en la señal.

El siguiente invariante a analizar es la dimensión de correlación. Para su estimación se implementaron los algoritmos propuestos por Grassberger y Procaccia, Judd y el algoritmo GKA, en las 3 estimaciones se utilizaron 19 valores diferentes para d_c (desde 2 hasta 20) y el valor estimado de $\tau_c = 4$.

Al observar los resultados de la aplicación del método de Grassberger Procaccia sobre la serie de la TRM (mostrados en la figura 19) estos parecen tener un área horizontal coincidente para las curvas entre los valores 0 y 2 en el eje de la abscisa. Si se calcula un promedio de los datos de todas las curvas alrededor de esta zona "horizontal" se tiene un valor para la dimensión de correlación de 0.9897 (Figura 20), que en primera instancia se puede pensar satisfactorio debido a que es muchísimo menor que la dimensión de embebimiento, y al ser menor que uno, hace pensar que el sistema subyacente está más cercano a un mapa discreto que a un sistema continuo.

Sin embargo, este cálculo tiene un gran inconveniente, y está relacionado con que el intervalo de valores en que se mueven las curvas en el rango es bastante amplio (desde alrededor de -2 hasta 5 en el eje de la ordenada) debido a esto resulta difícil hablar de un valor invariante de la dimensión de correlación.

Adicionalmente, en los resultados de las figuras 19 y 20, los valores de la dimensión de correlación d_c son significativamente menores que la dimensión de embebimiento que genera cada una de las curvas de la figura. De este modo el proceso no intenta llenar el espacio de embebimiento, lo cual significa que la dinámica predominante de la serie no es estocástica.

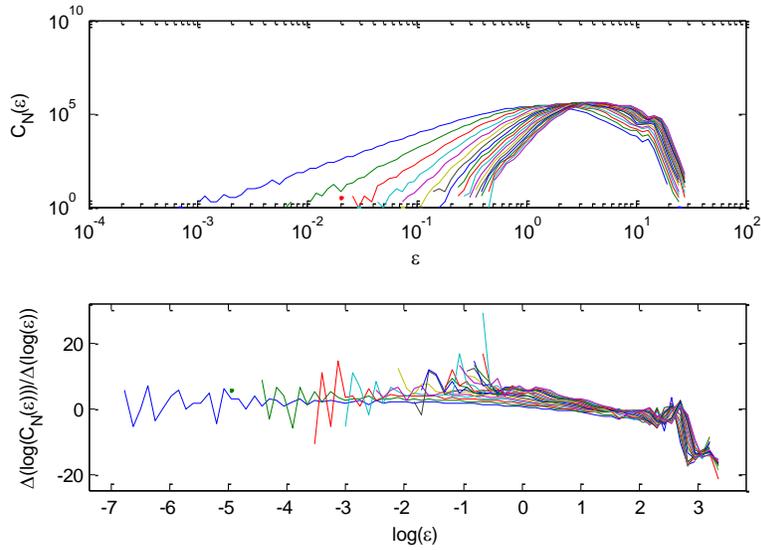


Figura 19: Calculo de la dimensión de correlación por medio del método de Grassberger y Procaccia

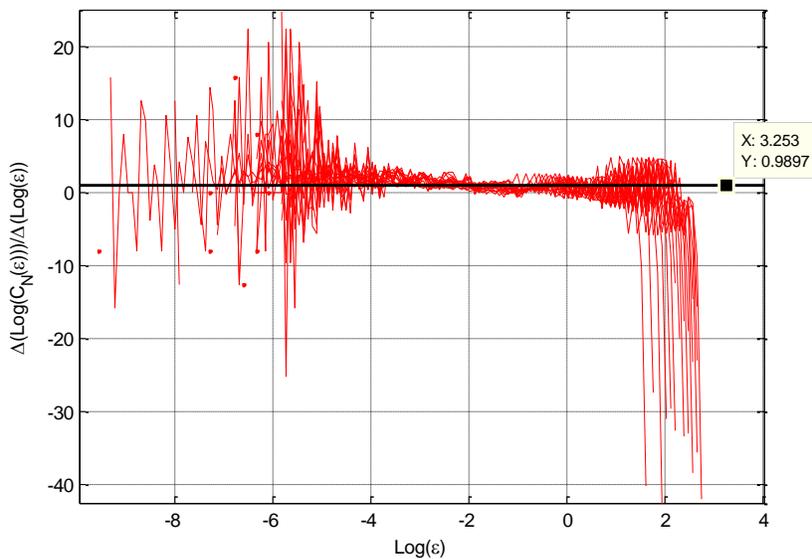


Figura 20: Calculo de la dimensión de correlación por medio del método de Grassberger y Procaccia

Los resultados entregados por el método de Judd (figura 21) en realidad no son muy concluyentes, ya que no es posible encontrar una zona “plana” común para todas las curvas. Esto solo confirma el problema hallado por medio del algoritmo de Grassberger – Procaccia con respecto a la amplia variabilidad de las curvas y la incapacidad para hallar un valor concreto de la dimensión de correlación.

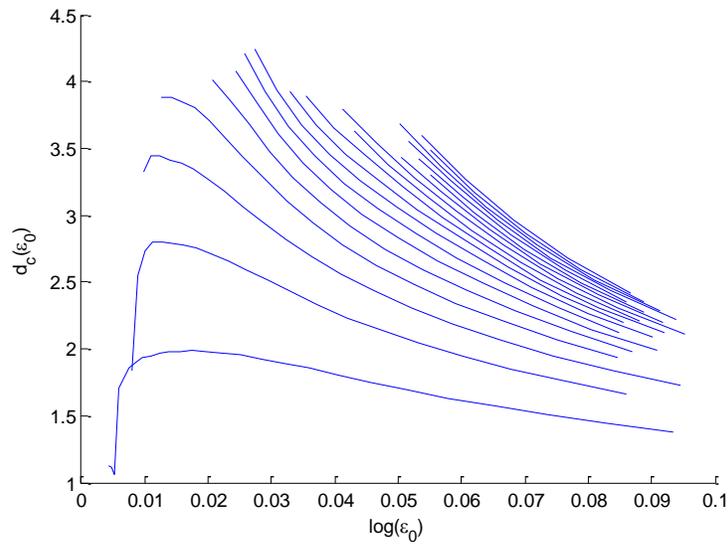


Figura 21: Calculo de la dimensión de correlación por medio del método de Judd

Los resultados entregados por el método GKA (figura 22) por su parte, muestran que el valor de la dimensión de correlación es cercano a 1 para las dimensiones menores que 8, sin embargo tampoco muestran una región horizontal en la que pueda deducirse un valor para la dimensión de correlación d_c . En este caso es posible argumentar que a pesar de que la curva de la dimensión de correlación es creciente los valores de d_c son mucho menores que su respectiva dimensión de embebimiento, luego la dinámica de la serie no satura el espacio de fase en ningún momento. En conclusión, aunque el método GKA no converge a un valor invariante para la dimensión de correlación demuestra una vez más que el origen de la series no es un proceso aleatorio,

Adicionalmente, el cálculo de la entropía, que es el tercer invariante definido en el capítulo 4, se realiza por medio del método GKA y se muestra también en la figura 22. El resultado muestra una curva con valores negativos, lo que claramente indica una falla en la convergencia del algoritmo para el cálculo de esta cantidad, ya que el valor mínimo de esta cantidad es cero.

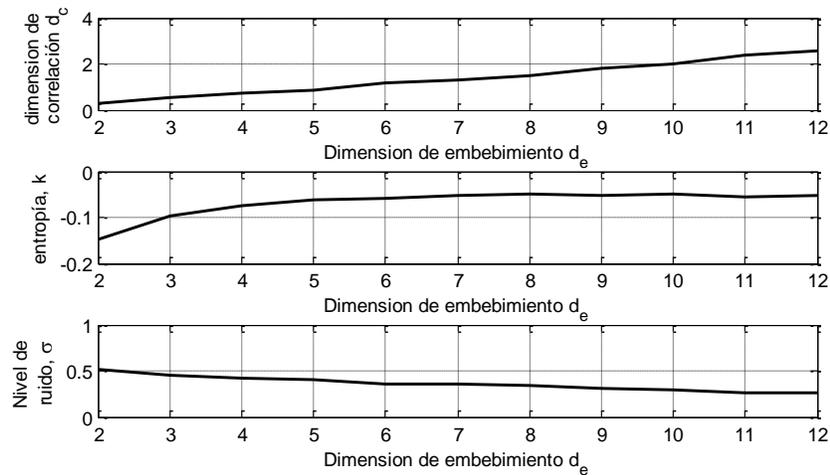


Figura 22: Cálculo de la dimensión de correlación, la entropía de la información y el nivel de ruido del sistema por medio del algoritmo GKA

Finalmente se tiene el cálculo del nivel de ruido en la serie, éste disminuye con el aumento de la dimensión de embebimiento como puede observarse en la figura 22. Esto puede ser debido a que al incrementar la dimensión de embebimiento el modelo intenta replicar parte de las dinámicas de la serie que en dimensiones menores se consideran ruido. El incremento desmedido de la dimensión de embebimiento puede ocasionar fenómenos de sobre-ajuste.

El último de los invariantes que queda por calcular es la entropía, ya que su estimación mediante el método GKA no fue posible. Es por esto que se calcula de nuevo mediante el método de Shannon. Este último método arrojó los resultados que se muestran en la figura 23. Como se puede observar, al comparar la entropía de la TRM con la de un proceso aleatorio y con un proceso totalmente determinista, como es la función seno, no existe mayor diferencia entre la entropía de la TRM y la del proceso aleatorio.

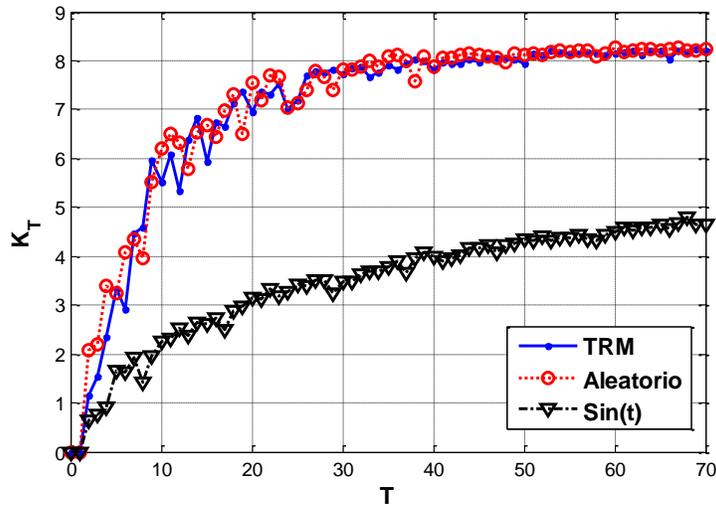


Figura 23: Cálculo de la dimensión de la entropía de la información por medio del método de Shannon

Cabe anotar que dentro del cálculo de invariantes, el error de predicción no lineal (NLPE) propuesto por (Small 2005) como cálculo alternativo al máximo exponente de Lyapunov depende principalmente del modelo seleccionado para el ajuste de las dinámicas del sistema subyacente. De esta forma y sabiendo que la dimensión de embebimiento debería tener un valor mínimo que a la vez garantice el despliegue efectivo del atractor, se puede plantear la posibilidad de utilizar la estimación de este error en la elección de los parámetros de embebimiento para la reconstrucción del espacio de fase del sistema, de forma que el valor del NLPE sea mínimo.

Calculando una matriz cuyas entradas son los resultados del NLPE para determinados valores de d_e y τ_e se puede elaborar el gráfico de contorno mostrado en la figura 24. Este diagrama puede ser de gran utilidad, al menos en la selección de valores pequeños de d_e , ya que aunque la grafica da a entender que mientras más grande sea el valor de la dimensión más bajo será el NLPE. Es posible que esta última afirmación sea cierta, pero es debido a que en dimensiones tan altas existe el riesgo de sobre-ajustar los datos y de esta forma el modelo intentará ajustar componentes que pueden ser consideradas como ruido.

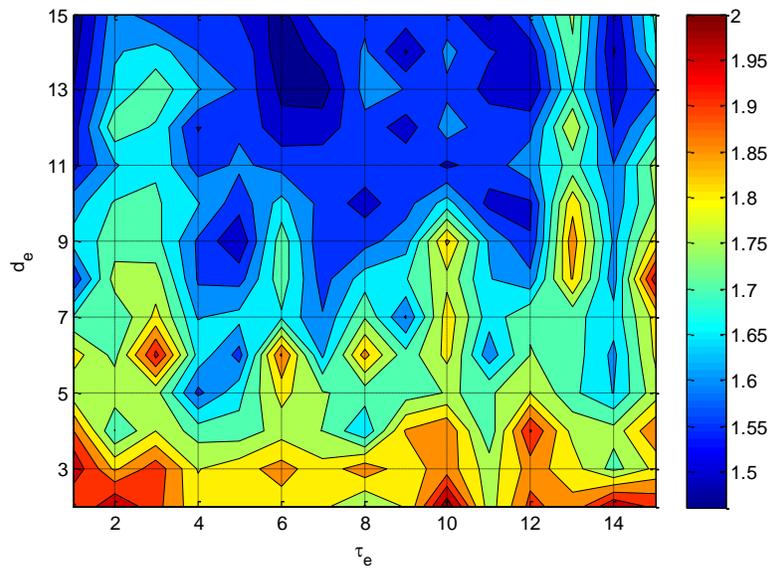


Figura 24: Cálculo del error no lineal de pronóstico para diferentes valores de τ_e y d_e

6.1.5 Análisis Lineal por medio de Surrogate Data

Construyendo los conjuntos de surrogate data a partir de la serie original de la TRM y con los tres algoritmos introducidos por Theiler sobre el análisis lineal con surrogate data, se puede proceder a aplicar las pruebas de hipótesis de Kolmogorov – Smirnov y la prueba de hipótesis χ^2 de Pearson para la bondad de ajuste entre distribuciones. La medida seleccionada para el análisis es la distribución de las distancias entre puntos usada para el cálculo de la dimensión de correlación d_c . Los resultados de las pruebas de hipótesis se muestran a continuación.

Prueba de Kolmogorov – Smirnov.

Algoritmo 0:

H_0 =Los datos son independientes e idénticamente distribuidos (iid).

Resultado: H_0 no puede ser rechazada.

Algoritmo 1:

H_0 =Los datos corresponden a un ruido linealmente filtrado, (un proceso ARMA)..

Resultado: H_0 no puede ser rechazada.

Algoritmo 2:

H_0 =Los datos corresponden a una observación no lineal de un ruido linealmente filtrado.

Resultado: H_0 no puede ser rechazada.

Prueba χ^2 de Pearson.

Algoritmo 0:

H_0 = No hay diferencia significativa entre las frecuencias de una distribución iid y las frecuencias de la distancia entre puntos calculada sobre la TRM.

Resultado: H_0 se rechaza.

Algoritmo 1:

H_0 = No hay diferencia significativa entre las frecuencias de la distribución de un ruido linealmente filtrado (un proceso ARMA) y las frecuencias de la distancia entre puntos calculada sobre la TRM.

Resultado: H_0 se rechaza.

Algoritmo 2:

H_0 = hay diferencia significativa entre las frecuencias de la distribución de una observación no lineal de un ruido linealmente filtrado y las frecuencias de la distancia entre puntos calculada sobre la TRM.

Resultado: H_0 se rechaza.

Dos situaciones pueden presentarse: 1) los datos tienen una gran componente estocástica en sus dinámicas, lo cual causa la inhabilidad para rechazar la hipótesis nula en la prueba de Kolmogorov – Smirnov; ó 2) los datos de hecho son el resultado de un sistema dinámico de dimensión muy alta, el cual no es lineal.

Por otro lado puede afirmarse que la serie de tiempo no es un proceso aleatorio, prueba de ello es la persistencia encontrada en el análisis de reescalamiento y el resultado del exponente de Hurst típicos de un proceso de memoria larga.

Si las dinámicas de la serie de la TRM partieran de un proceso ARMA, el sesgo causado por sus componentes autorregresivos y de media móvil se eliminaría en gran parte con el ajuste del modelo $AR(1)$, y es muy probable que en este caso el exponente de Hurst mostrara comportamiento aleatorio o incluso antipersistente (Peters 1994).

Conclusiones y discusión

Se realizó una revisión rigurosa de algunas herramientas usadas, a nivel mundial, para el análisis no lineal de series de tiempo en diferentes campos de la ingeniería y la física que han sido aplicadas recientemente al modelamiento financiero. Además se mostró un caso de aplicación de relevancia para el entorno colombiano como es el análisis de la serie de tiempo de la tasa representativa del mercado (TRM). Estas herramientas pueden ayudar a obtener una mejor interpretación y conocimiento de las dinámicas que gobiernan los movimientos erráticos presentes en series de tiempo en diversos campos de aplicación.

En el caso del análisis por medio del rango reescalado (R/S), muestra que el comportamiento de los rendimientos logarítmicos filtrados por medio de un proceso $AR(1)$ de la serie de tiempo de la TRM son persistentes con un alto grado de confianza, lo que quiere decir que la serie de datos presenta fenómenos de memoria en diferentes escalas de tiempo, especialmente de memoria larga. Esto es típico de las series de tiempo que presentan interdependencia no lineal o incluso caótica en su evolución. Este resultado alienta el desarrollo del análisis posterior.

A pesar de los resultados de las técnicas de análisis usadas (exponente de Hurst, función de información mutua, estimación del máximo exponente de Lyapunov, prueba χ^2 para el análisis de surrogate data) que sugieren la existencia de un tipo de dinámica que genera un efecto de memoria larga y, de la posibilidad de construir una emulación del espacio de fase del sistema dinámico de la TRM por medio de la teoría de embebimiento, la reconstrucción del espacio de fase no captura satisfactoriamente esta dinámica como se muestra en los resultados obtenidos a partir de los modelos de pronóstico y en el análisis de su espacio de fase por medio de la dimensión de correlación y la entropía de la información.

Se implementaron dos modelos de pronóstico a partir del embebimiento construido para la serie de la TRM. El primero fue modelo local constante (LCM) el cual se ajusta por medio de una regresión de mínimos cuadrados sobre los datos dispuestos en forma de matriz, donde cada fila constituye una dimensión del espacio de fase ajustado a la serie. El segundo modelo fue un perceptrón multicapa entrenado por medio de un algoritmo de regularización bayesiana donde las entradas al modelo son los rezagos del espacio de fase d_g – dimensional. Finalmente los resultados obtenidos de estos modelos fueron contrastados con el pronóstico por medio de un modelo $ARIMA(1,1,3)$. De los resultados obtenidos puede concluirse que ninguno de los métodos de pronósticos implementados sobre la TRM captura satisfactoriamente la información de esta

estructura ya que el error $RMSE$ que producen en el pronóstico a un paso es mayor que la variación diaria promedio de la variable.

Se presentó un posible uso del cálculo del error no lineal de pronóstico en la elección de los parámetros d_g y τ_g para la reconstrucción del espacio de fase del sistema. Esto teniendo en cuenta que para valores muy altos de la dimensión de embebimiento es posible que se presenten fenómenos de sobre-ajuste en el modelamiento de las dinámicas subyacentes que gobiernan el sistema, incluyendo posiblemente dinámicas estocásticas.

Por otro lado el cálculo de invariantes mostró resultados poco concluyentes sobre la presencia de caos determinista en las dinámicas de la serie de la TRM. Es posible que los datos presenten una gran cantidad de ruido, lo cual causa el fallo en la convergencia de los algoritmos a la hora de calcular la dimensión de correlación, la entropía de la información y la incapacidad de rechazar la hipótesis nula en la prueba de Kolmogorov-Smirnov. A pesar de esto, la prueba χ^2 de Pearson rechazó las hipótesis de que las distribuciones de las distancias entre puntos no tuvieran diferencia con una distribución de un proceso iid, una distribución de un proceso ARMA, y una distribución de observaciones no lineales de un proceso ARMA, es decir, los datos no proviene de un proceso lineal.

Como se dijo al inicio del trabajo, este introdujo algunas herramientas usadas para el análisis de no linealidad en series de tiempo. Sin embargo resultaba casi que imposible cubrir toda la variedad de herramientas desarrolladas alrededor de este tema. Es por esto que se propone como trabajo futuro el análisis de la serie de la TRM por medio de herramientas derivadas de la estadística fractal, como puede ser el análisis de ruido coloreado para una caracterización de la serie, y el ajuste de un proceso AFRIMA ó de una distribución de Levy, los cuales pueden replicar fenómenos de memoria larga.

Es probable que las herramientas propuestas como trabajos futuros presenten mejores resultados, debido a que los indicios recopilados en este trabajo indican que la TRM no está contaminada por una fuerte componente estocástica, sino que en realidad se comporta como un ruido fractal persistente ((Peters 1994) lo llama *Black Noise*) de alta dimensión en vez de un sistema dinámico determinista en régimen caótico que pueda representarse en un espacio de fase de baja dimensión. Esto podría explicar la falla de las herramientas que dependen de la reconstrucción del espacio de fase, como son el método de Judd y GKA para la estimación de la dimensión de correlación y la entropía de la información, además de los algoritmos de pronóstico como son el modelo local constante LCM y las redes neuronales.

Por otro lado, la convergencia del método de Cao y de la función de información mutua para la asignación de los valores d_g y τ_g seleccionados es también posible en el caso de que la señal sea un ruido fractal, ya que estas pueden ajustar componentes de baja dimensión presentes en la serie, pero que tal vez no son las más relevantes.

Bibliografía

- Abarbanel, Henry D.I. *Analysis of Observed Chaotic Data*. Springer Verlag, 1996.
- Broomhead, D.S., and Gregory P. King. "Extracting qualitative dynamics from experimental data." *Physical D* 20 (1985): 217 - 236.
- Burioka, Naoto, Germaine Cornélissen, Franz Halberg, and Daniel T. Kaplan. "Relationship between correlation dimension and indices of linear analysis in both respiratory movement and electroencephalogram." *Clinical Neurophysiology* 112 (2001): 1147-1153.
- Cao, Liangyue. "Practical method for determining the minimum embedding dimension of a scalar time series." *Phys. D* (Elsevier Science Publishers B. V.) 110 (1997): 43-50.
- Cao, Liangyue, and Abdol S. Soofi. "Nonlinear deterministic forecasting of daily dollar exchange rates." *International Journal of Forecasting* 15 (1999): 421-430.
- Carbajal, R., M.Vallverdú, R Baranowski, and P. Caminal. "Detección de determinismo no lineal en la variabilidad de la frecuencia cardiaca." *Memorias II Congreso Latinoamericano de Ingeniería Biomédica*, 2001.
- Cecen, A. Aydin, and Cahit Erkal. "Distinguishing between stochastic and deterministic behavior in high frequency foreign exchange rate returns: Can non-linear dynamics help forecasting?" *International Journal of Forecasting* 12 (1996): 465-473.
- De, Samuele G., Massimo Veltri, and Leonardo Primavera. "Estimated generalized dimensions of river networks." *Journal of Hydrology* 322 (2006): 181-191.
- Díaz, Andrés Fernández. *Dinámica caótica en Economía (Teoría y Aplicaciones)*. Mc Graw Hill, 2000.
- Eckmann, J. P., and D. Ruelle. "Ergodic theory of chaos and strange attractors." *Reviews of modern physics* 53 (1985): 617-656.
- Galka, A., T. Maaß, and G. Pfister. "Estimating the dimension of high-dimensional attractors: A comparison between two algorithms." *Physica D: Nonlinear Phenomena* 121 (1998): 237-251.
- Giraldo, Norman. "Predicción de betas y VAR De portafolios de acciones Mediante el filtro de Kalman y los modelos GARCH." *Cuadernos de Administración* 18 (2005): 103-119.
- Grassberger, Peter, and Itamal Procaccia. "Characterization of strange attractors." *Physical Review Letters*, 1983: 346 - 349.
- Jalil, Munir Andrés, and Martha Misas. "Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas." *Revista Colombiana de Estadística* 30 (2007): 143 a 161.

- Janjarasjitt, S., M.S. Scher, and K.A. Loparo. "Nonlinear dynamical analysis of the neonatal EEG time series: The relationship between neurodevelopment and complexity." *Clinical Neurophysiology* 119 (2008): 822-836.
- Jayawardena, A. W., and A. B. Gurung. "Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach." *Journal of Hydrology* 228 (2000): 242-264.
- Judd, Kevin. "An improved estimator of dimension and some comments on providing confidence intervals." *Physica D: Nonlinear Phenomena* 56 (1992): 216-228.
- Kantz, Holger, and Thomas Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 2006.
- Kohers, Theodor, Vivek Pandey, and Gerald Kohers. "Using nonlinear dynamics to test for market efficiency among the major U.S. stock exchanges." *The Quarterly Review of Economics and Finance* 37 (1997): 523-545.
- Kosmidis, Kosmas, Alkiviadis Kalampokis, and Panos Argyrakis. "Language time series analysis." *Physica A: Statistical Mechanics and its Applications* 370 (2006): 808-816.
- Lai, Ying-Cheng, and David Lerner. "Effective scaling regime for computing the correlation dimension from chaotic time series." *Physica D: Nonlinear Phenomena* 115 (1998): 1-18.
- Luo, Xiaodong. "Characterising Nonlinear Determinism From Experimental Time Series Data." The Hong Kong Polytechnic University, 2004.
- Mahajan, Arvind, and Andrew J. Wagner. "Nonlinear dynamics in foreign exchange rates." *Global Finance Journal* 10 (1999): 1-23.
- Matsuba, I., H. Suyari, S. Weon, and D. Sato. "Practical chaos time series analysis with financial applications." *5th International Conference on Signal Processing Proceedings* 1 (2000): 265-271.
- Mejia, Margarita. "Evaluación del modelo DENFIS para la predicción de series de tiempo económicas." Universidad Nacional de Colombia, Facultad de Minas, 2008.
- Misas, Martha, Enrique López, Juana Téllez, and José Fernando Escobar. "La inflación subyacente en Colombia: un enfoque de tendencias estocásticas comunes asociadas a un VEC estructural." *Lecturas de Economía* 63 (2005): 187-230.
- Mizrach, Bruce. "Determining delay times for phase space reconstruction with application to the FF/DM exchange rate." *Journal of Economic Behavior & Organization* 30 (1996): 369-381.
- Nakamura, Tomomichi, and Michael Small. "Testing for dynamics in the irregular fluctuations of financial data." *Physica A: Statistical Mechanics and its Applications* 366 (2006): 377-386.
- Onali, Enrico, and John Goddard. "Unifractal and multifractal in the Italian stock market." *International Review of Financial Analysis* 18 (2009): 154-163.

Ospina, Javier Humberto, and Edinson Caicedo Cerezo. "Un modelo estocástico sobre la predictibilidad del signo del retorno y su relación con la no linealidad en media." *Cuadernos de Administración* 21 (36) (2008): 11-35.

Ott, Edward, Tim Sauer, and James A. Yorke. *Coping with chaos : analysis of chaotic data and the exploitation of chaotic systems*. John Wiley and Sons, 1994.

Peters, Edgar E. *Fractal Market Analysis: Applying Chaos Theory to Investment & Economics*. John Wiley & Sons, 1994.

Saida, Ahmed Ben. "Using the Lyapunov Exponent as a Practical Test for Noisy Chaos." *SSRN eLibrary*, 2007.

Santana, Juan Camilo, Álvaro Andrés Camaro, Arnoldo Casas Henao, and Édgar Jiménez Méndez. "Estudio empírico sobre la capacidad predictiva de las redes neuronales en el pronóstico de la inflación colombiana: una metodología alternativa." *Revista INNOVAR*. 16 (2006): 1-12.

Small, Michael. *Applied Nonlinear Time Series Analysis: Applications In Physics, Physiology And Finance*. Edited by Berkeley Leon O. Vol. 52. World Scientific Publishing Co. Pte. Ltd., 2005.

Small, Michael, et al. "Deterministic nonlinearity in ventricular fibrillation." *Chaos: An Interdisciplinary Journal of Nonlinear Science (AIP)* 10 (2000): 268-277.

Soofi, Abdol S., and Liangyue Cao. "Nonlinear deterministic forecasting of daily Peseta-Dollar exchange rate." *Economics Letters* 62 (1999): 175-180.

Sprott, Julien Clinton. *Chaos and Time-Series Analysis*. Oxford University Press, 2003.

Strogatz, Steven H. *Nonlinear Dynamics and Chaos*. Perseus Publishing, 2000.

Strozzi, F., and J.M. Zaldívar. "Non-linear forecasting in high-frequency financial time series." *Physica A: Statistical Mechanics and its Applications* 353 (2005): 463-479.

Szpiro, George G. "Noise in unspecified, non-linear time series." *Journal of Econometrics* 78 (1997): 229-255.

Theiler, James. "On the evidence for low-dimensional chaos in an epileptic electroencephalogram." *Physics Letters A* 196 (1995): 335-341.

Theiler, James, Stephen Eubank, André Longtin, Bryan Galdrikian, and J. Doyne Farmer. "Testing for nonlinearity in time series: the method of surrogate data." *Physica D: Nonlinear Phenomena* 58 (1992): 77-94.

Velásquez, Juan David, and Santiago Montoya. "Modelado del índice de precios al consumidor usando un modelo híbrido basado en redes neuronales artificiales." *Dyna* 147 (2005): 85-93.

Velásquez, Juan David, Isaac Dyner, and Reinaldo C. Souza. "Tendencias en la predicción y estimación de los intervalos de confianza usando modelos de redes neuronales aplicados a series temporales." *Dyna* 149 (2006): 141-147.

Vélez, Elkin Castaño, and Luis Fernando Melo. "MÉTODOS DE COMBINACIÓN DE PRONÓSTICOS : Una aplicación a la inflación colombiana." BANCO DE LA REPUBLICA, Subgerencia de Estudios Económicos, 1999.

Wolf, Alan, Jack B. Swift, Harry L. Swinney, and John A. Vastano. "Determining Lyapunov exponents from a time series." *Physica D: Nonlinear Phenomena* 16 (1985): 285-317.

Yu, Dejin, et al. "Measuring temporal complexity of ventricular fibrillation." *Physics Letters A* 265 (2000): 68-75.

Zunino, Luciano, Alejandra Figliola, Benjamin M. Tabak, Darío G. Pérez, Mario Garavaglia, and Osvaldo A. Rosso. "Multifractal structure in Latin-American market indices." *Chaos, Solitons & Fractals* 41 (2009): 2331-2340.