PRUEBAS DE NO LINEALIDAD: EL MÉTODO DE LOS DATOS SUSTITUTOS.

Non-linearity test: The surrogate data method.

DIEGO L. GUARIN
Ingeniero Físico
Estudiante Maestría Ing. Eléctrica
Grupo de Investigación en Control e Instrumentación
Universidad Tecnológica de Pereira
dlguarin@gmail.com

CRISTIAN H. RODRIGUEZ
Estudiante Ingeniería Eléctrica
Universidad Tecnológica de Pereira
ingecristian99@gmail.com

ALVARO A. OROZCO
Ingeniero Eléctrico, Ph.D.
Docente Titular
Grupo de Investigación en Control e Instrumentación (Director)
Universidad Tecnológica de Pereira
aaog@utp.edu.co

RESUMEN
En este articulo se presenta una introducción al método de los datos sustitutos, se empieza por mencionar el procedimiento general de Monte Carlo para la prueba de hipótesis nulas, posteriormente se introduce el método de los datos sustitutos para pruebas de no linealidad, proponiendo una jerarquía de hipótesis nulas y una batería de estadísticas no lineales que permitan comparar el comportamiento de una serie real contra un conjunto de sustitutos generados de tal manera que satisfagan las hipótesis nulas, se presenta un criterio discriminante por medio del cual se podrá rechazar o aceptar la hipótesis. Finalmente se presenta un ejemplo del funcionamiento de los algoritmos utilizando la serie de Lorenz.

PALABRAS CLAVES: Dinámica no lineal, datos sustitutos, prueba hipótesis nulas, teoría del caos, estadística paramétrica y no paramétrica.

ABSTRACT
In the following we present an introduction to the method of surrogate data, we start by mentioning the general proceeding of the Monte Carlo hypothesis test, then we introduce the method of surrogate data for non linearity test, proposing an hierarchy of null hypothesis and a battery of non linear statistics that allow us to compare the behavior of a real time series against a set of surrogates which were generated to fit the null hypothesis, we present a discriminating criterion whereby we can accept o reject a null hypothesis. Finally we present an example of how the algorithm works using the Lorenz time series.

KEYWORDS: *Non linear dynamics, surrogate data, null hypothesis testing, chaos theory, parametric and non parametric statistics.*

1. INTRODUCCIÓN

Existen dos razones que motivan la aplicación de métodos no lineales cuando se analiza una serie temporal. La primera es que a pesar de haber utilizado un conjunto de métodos lineales algunas estructuras de la serie de tiempo aún son desconocidas. En cuanto a la segunda, es posible que se tenga un conocimiento previo del comportamiento no lineal del sistema intrínseco, por lo tanto una descripción lineal parece inadecuada. Este argumento es comúnmente escuchado cuando se analizan series biológicas, por ejemplo, no se espera que el cerebro tenga un comportamiento lineal [1]. Sin embargo, este tipo de razonamiento es peligroso. El hecho que un sistema dinámico contenga componentes no lineales, no implica que estos estén reflejados en la señal específica que se ha medido. Tampoco se conoce si tiene interés practico ir más allá de la aproximación lineal. Al final de cuentas, no se desea que el análisis realizado sobre la serie temporal refleje los prejuicios del investigador sobre el sistema dinámico subyacente, sino que se realice de acuerdo a las estructuras presentes en los datos. Por lo tanto, previo a la aplicación de cualquier método de análisis no lineal a la serie de datos, se debe determinar si esta es de hecho no lineal.

Para alcanzar este objetivo Theiler et al. [2] desarrollaron el método de los datos sustitutos, el cual se emplea para clarificar y cuantificar declaraciones sobre la presencia de no linealidad en series temporales. Este método consiste básicamente en comprar el valor de una estadística no lineal para los datos y para varias clases de sistema lineales, de tal manera que se pueda probar si los datos tienen alguna característica diferente a los sistemas estocásticos lineales. Es común [2-6] utilizar estadísticas basadas en la dimensión de correlación [7] para probar hipótesis que los datos provienen de alguna clase de sistema dinámico lineal.

A continuación se realiza una introducción al formalismo del método de Monte Carlo para probar hipótesis y a las terminologías utilizadas en el método de los datos sustitutos. Posteriormente se introduce una jerarquía de hipótesis nulas y los correspondientes algoritmos para generar los sustitutos basados en dichas hipótesis, también se proponen dos formas de verificar la aceptación o rechazo de dichas hipótesis. Finalmente se presenta un ejemplo de la aplicación de los algoritmos a la serie temporal de Lorenz.

Fecha de Recepción: Enero 26 de 2010
Fecha de Aceptación: Marzo 25 de 2010

2. MÉTODO DE MONTE CARLO PARA PROBAR HIPÓTESIS

El test de hipótesis nula usa medidas estadísticas del sistema subyacente para determinar la probabilidad que una hipótesis propuesta sea verdadera (o falsa). El procedimiento a seguir suele ser [8]:

- Se formula la hipótesis nula de interés, la hipótesis alternativa y los riesgos potenciales asociados con cada decisión.
- Se eligen las pruebas estadísticas.
- Calculo de la distribución de frecuencias de la prueba estadística bajo la hipótesis nula.
- Con la guía de la distribución de frecuencias, se elige un criterio para determinar si se rechaza la hipótesis o no.

La idea básica es producir diferentes realizaciones de acuerdo con la hipótesis nula a través de las simulaciones de Monte Carlo, estas realizaciones son usualmente generadas a partir de los datos experimentales y son llamadas sustitutos. Del conjunto de sustitutos se calcula una distribución empírica y el intervalo de confianza de la prueba estadística. Dado esto, la distribución de frecuencia depende esencialmente del algoritmo por medio del cual se generaron los sustitutos

2.1 Realización articular y limitada de sustitutos

Siguiendo la notación introducida en [3] sea $\boldsymbol{\mathcal{F}}$ el grupo de todos los posibles procesos para el problema bajo consideración. Sea $\emptyset$ la hipótesis nula y $\boldsymbol{\mathcal{F}}_{\emptyset}$ el conjunto de procesos que son consistentes con la hipótesis $\emptyset$. Si $\boldsymbol{\mathcal{F}}_{\emptyset}$ consiste en un solo elemento entonces la hipótesis nula es llamada simple, de lo contrario es llamada compuesta.

Dada una hipótesis nula compuesta $\emptyset$ y un proceso $\boldsymbol{F}$ consistente con ella, se puede denotar la estadística elegida como $\boldsymbol{T}$, y la correspondiente función de distribución de amplitudes (a.d.f) correspondiente a la hipótesis nula por $\boldsymbol{P_{T,F}(t) \equiv Prob(T < t|F \in \mathcal{F}_{\phi})}$. Si para cualquier dos procesos $\boldsymbol{F_i}$ y $\boldsymbol{F_j}$ $(\boldsymbol{i \neq j})$ en el conjunto $\boldsymbol{\mathcal{F}}_{\emptyset}$, se tiene que $\boldsymbol{P_{T,F_i} = P_{T,F_j}}$ entonces se dice que la estadística $\boldsymbol{T}$ es articular, de lo contrario es no articular. Una gran ventaja de las estadísticas articulares, es que la distribución $\boldsymbol{P_{T,F}(t)}$ es independiente del proceso $\boldsymbol{F}$. Suponga que la serie $\boldsymbol{d = \{d_i\}_{i=1}^{n}}$ bajo prueba es generada por un proceso $\boldsymbol{F_d \in \mathcal{F}_{\phi}}$, para evitar falsos rechazos de la hipótesis nula, es una condición suficiente que el proceso $\boldsymbol{F_s}$ que produce los datos sustitutos satisfaga $\boldsymbol{F_S = F_d}$. Los sustitutos generados por dicho proceso son llamados limitados, de lo contrario son llamados no limitados. Dado un conjunto de sustitutos limitados $\{\boldsymbol{s_1}, \boldsymbol{s_2}, \dots\}$, cualquier estadística $\boldsymbol{T}$ se muestra como si fuera articulada para los procesos $\{\boldsymbol{F_d}, \boldsymbol{F_{s_1}}, \boldsymbol{F_{s_2}}, \dots\}$.

Dada la importancia que tiene el algoritmo por medio del cual se generan los sustitutos el método de Monte Carlo para probar hipótesis es comúnmente llamado prueba de datos sustitutos, o método de los datos sustitutos [2,3,9], y una de sus aplicaciones más importantes como ya se había mencionado es la detección de no linealidad en series de tiempo [2,10-12].

3. DATOS SUSTITUTOS PARA DETECCIÓN DE NO LINEALIDAD

A la hora de detectar la presencia de no linealidades en una serie de tiempo es posible utilizar una estrategia directa, es decir se puede calcular algunas características no lineales tales como la dimensión de correlación, el espectro de Lyapunov [13,14], entre otras, como medidas discriminantes con la creencia que estas estadísticas revelan la presencia de no linealidad en el sistema [15]. En la práctica esta estrategia puede arrojar resultados adversos, ya que se ha demostrado que en ciertas situaciones las características no lineales no son inequívocas del sistema subyacente [3], algunos procesos estocásticos lineales también tienen dimensión fractal finita como la mayoría de los sistemas no lineales [16].

Una estrategia alternativa es aplicar el método de datos sustitutos. Para la aplicación del método primero se debe proponer una hipótesis nula la cual generalmente asume que la serie temporal es generada por un sistema estocástico lineal. Dada la hipótesis se produce un conjunto de sustitutos a partir de la serie original. Posteriormente se elige una estadística adecuada. Una vez calculada la estadística de prueba, se verifica si el valor de la estadística de los datos originales es típico de la distribución que presenta la estadística de los sustitutos de acuerdo a cierto criterio. Si esto no ocurre, se rechaza la hipótesis con cierto nivel de confianza (el cual depende de la estadística).

A continuación se expone la jerarquía de hipótesis propuestas por Theiler et al. [2], también se presentando un conjunto de algoritmos introducidos en [4], los cuales generan sustitutos que satisfagan las hipótesis pero superando algunos de los problemas que existían originalmente.

3.1 Jerarquía de pruebas de datos sustitutos

La siguiente discusión está limitada a series temporales estacionarias irregulares. Dado que los datos estacionarios generados por un proceso lineal determinista aparecen como periódicos o constantes, es trivial identificar entre lineal estocástico y determinista (siempre que el ruido no sea muy grande), por lo tanto solo se consideran los casos en que la serie irregular es generada por un sistema lineal estocástico o por un sistema no lineal estocástico o determinista.

3.1.1 Hipótesis nulas

La asunción básica es que la serie bajo prueba proviene de un proceso lineal estocástico ruidoso, ya sea i.i.d (distribuido de manera idéntica e independiente) o correlacionado (en general de un modelo auto regresivo de media móvil (ARMA)). Si durante la medición se permite que los datos originales pasen a través de un filtro no lineal entonces se introduce un componente no lineal a los datos. Teniendo esto en consideración se puede formular la siguiente jerarquía de hipótesis nulas, como se muestra en la figura 1 (i.e, la hipótesis NH0 es un caso especial de la hipótesis NH1, y así sucesivamente).

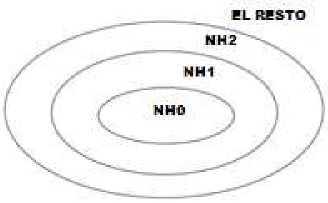

Figura 1. Representación esquemática de la jerarquía de hipótesis

- **Hipótesis Nula 0 (HN0):** Los datos de prueba son ruido i.i.d con media y varianza desconocidos.
- **Hipótesis Nula 1 (HN1):** Los datos de prueba son producidos por un proceso lineal estocástico en la forma de un modelo ARMA con parámetros desconocidos, lo cual es un filtro lineal del ruido i.i.d.
- **Hipótesis Nula 2 (HN2):** Los datos de prueba se obtienen al aplicar un filtro no lineal estático y monotónico a la serie original generada por un proceso ARMA**.**

3.2 Pruebas estadísticas

En principio la única preocupación que debe existir con relación a la estadística es que las mediciones hechas a los datos originales sean consistentes con los sustitutos si la hipótesis resulta ser cierta, de lo contrario mostrara discrepancia. Como se mencionó una elección popular es la dimensión de correlación, dado que ha sido demostrado que esta es una estadística articulada para la jerarquía de hipótesis nulas para la detección de no linealidad [3]. Existen otro conjunto de candidatos posibles, en [2] se demuestra que algunas estadísticas pueden entregar falsos rechazos o falsas aceptaciones, por lo tanto es importante poseer una batería de pruebas. Además de la dimensión de correlación algunas de las estadísticas más populares son los exponentes de Lyapunov, el error de predicción y la medición de la asimetría al invertir el tiempo [2, 3, 4, 9, 17].

3.3 Algoritmos para generar sustitutos

A continuación se introducen los algoritmos para generar sustitutos no paramétricos y limitados que corresponden con la jerarquía de hipótesis nula mencionada anteriormente.

3.3.1 Algoritmo 0

El algoritmo 0 prueba la hipótesis que los datos se comportan como ruido i.i.d, por lo tanto los sustitutos no deben ser correlacionados pero deben tener una distribución de probabilidad idéntica a la de los datos originales.

Algoritmo 0: versión 1 El sustituto $\boldsymbol{S_i}$ es creado al aleatorizar el orden de los datos $\boldsymbol{Z}$. Se genera un grupo de datos i.i.d $\boldsymbol{Y}$ y se reordena $\boldsymbol{Z}$ de tal forma que tenga la misma distribución de rangos que $\boldsymbol{Y}$. El sustituto $\boldsymbol{S_i}$ es el reordenamiento de $\boldsymbol{Z}$.

El procedimiento mencionado es llamado muestreo sin reemplazo. Esto significa que los datos y los sustitutos tendrán exactamente la misma distribución de probabilidades. Pero esto es más de lo que se espera de diferentes observaciones de un sistema estocástico. Para superar esta situación se propone una modificación del algoritmo llamada muestreo con remplazo.

Algoritmo 0: versión 2 El sustituto $\boldsymbol{S_i}$ es creado al muestrear de manera aleatoria con reemplazo la serie $\boldsymbol{Z}$.

Con esta versión la distribución de $\boldsymbol{S_i}$ es un aproximado de la de $\boldsymbol{Z}$. En la práctica de ha notado que para series con poca cantidad de datos la versión 1 funciona mejor, mientras que cuando se tiene una gran cantidad de datos la versión 2 arroja mejores resultados [3].

3.3.2 Algoritmo 1

Este algoritmo prueba la hipótesis que los datos se comportan como ruido filtrado. Este tipo de señales se caracteriza por su espectro de potencia. El algoritmo 1 genera sustitutos aleatorizando las fases de la transformada de Fourier de los datos.

Algoritmo 1: versión 1 Los sustitutos $\boldsymbol{S_i}$ con este algoritmo se producen al aplicar el algoritmo 0 a la fase de la transformada de Fourier de $\boldsymbol{Z}$. Se calcula la inversa de la transformada de Fourier para obtener los sustitutos $\boldsymbol{S_i}$.

Este algoritmo hereda los problemas que tiene el algoritmo 0. Una versión más refinada del mismo es la siguiente.

Algoritmo 1: versión 2 En vez de aplicar el algoritmo 0 a la fase de la transformada de Fourier de $\boldsymbol{Z}$, esta se multiplica por un factor $\boldsymbol{e^{i\phi}}$, donde $\boldsymbol{\phi}$ es un número aleatorio distribuido de manera uniforme en el intervalo $[\boldsymbol{0}, \boldsymbol{2\pi})$. Se toma la inversa de la transformada de Fourier para producir los sustitutos $\boldsymbol{S_i}$.

Aun ejecutando el procedimiento de esta manera es posible obtener algunos sustitutos con el mismo espectro de potencias que los datos originales. Sin embargo, aun no es claro como adherir un poco de aleatoriedad a los sustitutos sin sobre aleatorizarlos.

3.3.3 Algoritmo 2

Este algoritmo prueba la hipótesis que los datos se comportan como una transformación no lineal monotónica de ruido lineal filtrado. Esto se puede probar de dos maneras, la primera es re-escalar los datos de tal forma que sean Gaussianos y posteriormente se aplica el algoritmo 1. También se puede alcanzar el mismo objetivo utilizando el siguiente algoritmo.

Algoritmo 2: Se empieza con los datos $\boldsymbol{Z}$, se genera un grupo de datos con distribución gaussiana $\boldsymbol{Y}$ y se reordena $\boldsymbol{Y}$ de tal manera que tenga la misma distribución de rangos que $\boldsymbol{Z}$. Posteriorme se crean unos sustitutos $\boldsymbol{S_i}$ de $\boldsymbol{Y}$ usando el algoritmo 1: versión 2. Finalmente se reordena la serie original $\boldsymbol{Z}$ para crear unos sustitutos $\boldsymbol{S_i}$ los cuales tiene la misma distribución de rangos que $\boldsymbol{Y}$.

Los sustitutos que se generan de esta forma son conocidos como sustitutos con amplitud ajustada de la transformada de Fourier (AAFT). En este caso no se tiene los problemas que existen con los algoritmos anteriores, los datos obtenidos no son demasiado similares a los originales.

Este método posee una gran cantidad de problemas. A pesar que $\boldsymbol{Z}$ y $\boldsymbol{S_i}$ tienen distribuciones de probabilidad idénticas, en general, no tendrán un espectro de Fourier idéntico. Para resolver esto, Schreiber y Schmitz [18] sugirieron el siguiente procedimiento llamado IAAFT: Después de re-escalar los datos para preservar la distribución de probabilidad, la aleatorización de las fases altera la distribución de los rangos. Por lo tanto se debe re-escalar las fases. Pero, esto reduce la aleatoridad en las fases de la transformada de Fourier, entonces, es necesario re-aleatorizar las fases. Este procedimiento se repite hasta que se obtenga una convergencia razonable tanto en la distribución de los rangos y el espectro de potencia.

Desafortunadamente, este método no garantiza convergencia en el espectro de Fourier, pero los resultados muestran que siempre se llega a cierto grado de igualdad entre los espectros de potencia [3]. Esta solución introduce el mismo problema que se tenía con los algoritmos anteriores, los sustitutos se pueden volver sobre-limitados, esto depende de la cantidad de iteraciones del algoritmo AAFT (IAAFT).

Otro problema es que, para aplicar la transformada de Fourier, los datos se deben asumir como periódicos, una posible solución sugerida en [9] es llevar a cabo una fase de aleatorización limitada. Otra posible solución es no adoptar la transformada de Fourier, como se discute en [3]. Finalmente los problemas relacionados con los efectos de borde se observan en los procesos que involucran la transformada de Fourier (algoritmos 1 y 2), si existe una gran discrepancia entre el inicio y el fin de los datos, la transformada de Fourier lo considera como una discontinuidad en el sistema subyacente. La mejor solución es seleccionar una sub serie a partir de la serie original la cual posea una discrepancia mínima entre el inicio y el final, y con esta sub serie generar los sustitutos [18].

3.4 Criterio discriminante

Dado que el conocimiento exacto de la distribución estadística no siempre está disponible, se debe usar cierto criterio discriminante que permita tomar una decisión con cierto nivel de confianza. Existen dos clases de criterios: paramétricos y no paramétricos. El criterio paramétrico asume que la estadística sigue una distribución Gaussiana, y los parámetros de dicha distribución (media y varianza) se estiman de las muestras finitas. Se puede determinar cuándo rechazar o aceptar la hipótesis examinando si la estadística de la serie original sigue la distribución estadística de los sustitutos, el correspondiente nivel de confianza se puede calcular de la distribución estadística. El criterio no paramétrico [19] examina los rangos de las valores de la estadística de los datos originales y de los sustitutos. Suponga que la estadística de la serie original es $\boldsymbol{T_0}$ y el valor para los sustitutos es $\{\boldsymbol{T_i}\}_{i=1}^{N}$ dadas $\boldsymbol{N}$ realizaciones de sustitutos. Entonces si la estadística de la serie temporal original y de los sustitutos siguen la misma distribución, la probabilidad de que $\boldsymbol{T_0}$ sea menor o mayor que el conjunto $\{\boldsymbol{T_i}\}_{i=1}^{N}$ esta dada por $\mathbf{1}/(\boldsymbol{N}+\mathbf{1})$. Por lo tanto con $\boldsymbol{N}$ grande, cuando se encuentra que $\boldsymbol{T_0}$ es mayor o menor que cualquier valor en $\{\boldsymbol{T_i}\}_{i=1}^{N}$es muy probable que $\boldsymbol{T_0}$ siga una distribución diferente que $\{\boldsymbol{T_i}\}_{i=1}^{N}$. Por lo tanto el criterio rechaza una hipótesis nula si la estadística $\boldsymbol{T_0}$ es la mayor o menor en el conjunto $\{\boldsymbol{T_0}, \boldsymbol{T_1}, \dots, \boldsymbol{T_N}\}$, el nivel de confianza es $\mathbf{1}/(\boldsymbol{N}+\mathbf{1})$ para pruebas de una sola cara y $\mathbf{2}/(\boldsymbol{N}+\mathbf{1})$ para pruebas de dos caras.

4. EJEMPLO: SERIE DE LORENZ

El sistema de Lorenz [20] fue el primer sistema en ser reconocido como un sistema caótico, se describe por medio de (1) y exhibe caos cuando $\boldsymbol{\alpha}=\mathbf{16}$, $\boldsymbol{\beta}=\mathbf{45.92}$, $\boldsymbol{\gamma}=\mathbf{4}$. Se encontró que la dimensión de embebimiento y el tiempo de retardo son 3 y 4 respectivamente [21].

$$\frac{\partial x}{\partial t}=\alpha(y-x),\ \frac{\partial y}{\partial t}=x(\beta-z-y),\ \frac{\partial z}{\partial t}=xy-\gamma z \quad (1)$$

Para el siguiente ejemplo se solucionó el sistema de ecuaciones con la función ode45 de Matlab usando una frecuencia de muestreo de 25Hz, se toma la serie $\boldsymbol{x}$ como la serie temporal a analizar. Buscando tener un nivel de confianza del 95% se calcularon19 sustitutos con cada algoritmo, es decir se calcularon un total de 57 sustitutos.

En la Figura 2 se presenta la serie original y uno de los sustitutos generados con cada uno de los algoritmos, se observa que la serie no se asimila a ninguno de los sustitutos, por lo que se puede sospechar que ninguna de las hipótesis nulas es cierta.

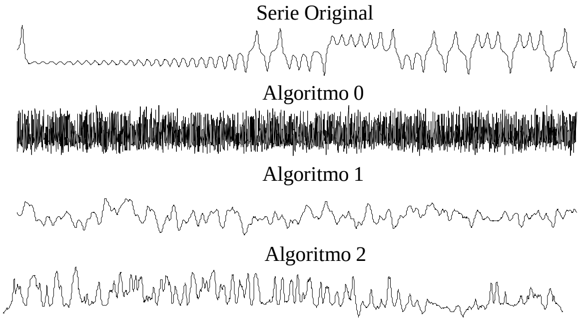

Figura 2. Serie de datos original y un sustituto generado con cada algoritmo.

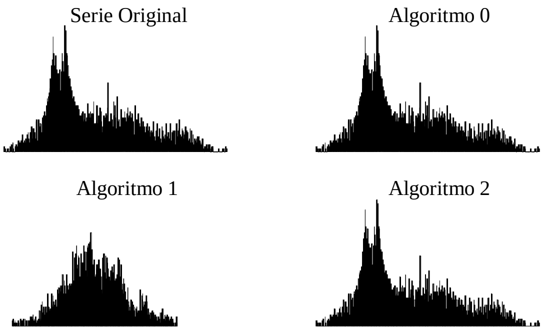

Figura 3. a.d.f para la serie original y para los sustitutos generados con cada uno de los algoritmos.

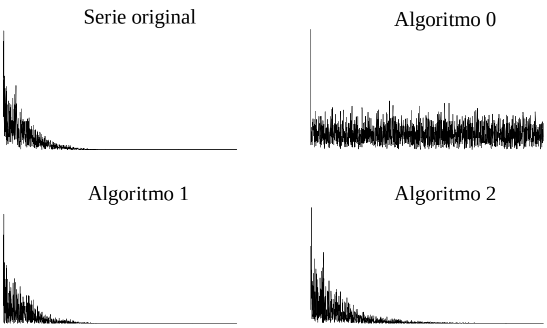

Figura 4. Espectro de potencia para la serie original y para los sustitutos generados con cada uno de los algoritmos.

Las Figuras 3 y 4 muestran respectivamente la distribución de probabilidad y el espectro de potencias para la serie original y para los sustitutos. Se observa que el algoritmo 0 preserva la a.d.f pero altera el espectro de frecuencias, convirtiéndolo en un espectro continuo como se espera del ruido i.i.d, el algoritmo 1 preserva el espectro de frecuencias pero no la a.d.f la cual se torna Gaussiana, mientras que el algoritmo 2 conserva tanto la a.d.f como el espectro de frecuencias. También se puede notar que tanto las a.d.f como los espectros de frecuencias de los sustitutos no son exactos al de la serie original, pero son lo suficientemente similares, para obtener un espectro idéntico se requiere la aplicación del algoritmo IAAFT.

5. DISCUSIÓN

Uno de los mayores problemas que se presenta con la aplicación del método es su utilización por parte de los investigadores como un algoritmo de caja negra, está demostrado que en ciertas situaciones algunas estadísticas entregan falsas aceptaciones o falsos rechazos de las hipótesis nulas [2], por lo tanto es necesario tener todo un arsenal de estadísticas para validar o rechazar las hipótesis, en la literatura se pueden encontrar 8 estadísticas no lineales que entregan buenos resultados [22], pero por lo general se usan solo 3.

Pero el hecho de utilizar una estadística particular implica seleccionar correctamente los parámetros inherentes a ella, por ejemplo se ha mostrado que el número de vecinos utilizado a la hora de calcular el error de predicción influye en el porcentaje de falsas aceptaciones o falsos rechazos [19].

Dados los problemas que se han presentado a la hora de aplicar los algoritmos a series reales, existe en la actualidad un interés por desarrollar nuevas maneras de generar sustitutos que no presenten los problemas que poseen los anteriores algoritmo, uno de los métodos más prometedores es el presentado por M. Small [3] el cual se basa en la idea que los sistemas lineales cumplen el principio de superposición, es decir si una serie $\boldsymbol{Z_1}$ y otra $\boldsymbol{Z_2}$son generadas por un proceso lineal, entonces la serie $\boldsymbol{Z_1 + Z_2}$ también lo será, por el contario si ambas son generadas por un proceso no lineal entonces la suma de estas será más complicada que cada una de las que lo forman. Este algoritmo está aun en etapa de pruebas pero sus resultados previos parecen prometedores.

Otro aspecto importante es que las hipótesis presentadas únicamente verifican si el sistema subyacente se comporta como un sistema lineal, es interesante comparar la serie contra hipótesis no lineales, como se propone en [9].

6. CONCLUSIONES

Se presentó un método por medio del cual es posible comparar una serie de datos contra un conjunto de hipótesis nulas las cuales intentan modelar el sistema dinámico que génere la serie. Estas hipótesis intentan verificar que el sistema dinámico es un sistema gaussiano lineal, si todas las hipótesis son rechazadas entonces existe la posibilidad que el sistema dinámico sea no lineal (caótico determinista, caótico no determinista o no lineal estocástico) o simplemente un sistema dinámico lineal que se comporta de una manera diferente a las que se modelaron con las hipótesis. Es importante recalcar entonces que el hecho que sean rechazadas las hipótesis nulas no implica que el sistema sea caótico determinista como ha sido erróneamente asumido por algunos autores.

Se presentaron los algoritmos por medio de los cuales es posible generar sustitutos consistentes con cada hipótesis, se mencionaron los problemas de cada algoritmo y se propusieron diversas soluciones, aunque estos algoritmos son aun un problema abierto y en la actualidad se busca desarrollar nuevos que eviten ciertos problemas.

También se hizo mención de las diferentes estadísticas disponibles para rechazar o aceptar las hipótesis y los criterios discriminantes para realizar esta tarea, se hizo especial énfasis en la necesidad de tener una batería de estadísticas ya que en ciertas situaciones particulares una estadística puede dar un falso rechazo o una falsa aceptación de una hipótesis**.**

Agradecimientos: Diego L. Guarín desea agradecer a la red de universidades de eje cafetero Alma Mater que ha financiado su investigación bajo el código AMFRI - JI08-09-01.

7. BIBLIOGRAFIA

[1] H. Kantz, J. Kurths, and G. Mayer-Kress, *Non Linear Analysis of Physiological data*, Berlin: Springer, 1998.

[2] J. Theiler, S. Eubank, A. Longtin, B. Galdrikan, and J. D. Frame, "Testing for nonlinearity in time series: The method of surrogate data," *Physica D,* vol 58, pp. 77-94, 1992.

[3] M. Small, *Applied Non Linear Time Series Analysis Applications in Physics, Physiology and Finance,* Singapore: World Scientific Publishing, 2005.

[4] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D,* vol 142, pp. 346-382, 2000.

[5] M. Small and K. Judd, "Correlation dimension: A pivotal statistic non-constrained realizations of composite hypotheses in surrogate data analysis," *Physica D*, vol 120, pp.386-400, 1998.

[6] A. I. Mees, *Non Linear Dynamics and Statistics*, USA: Birkhauser, 2001.

[7] P. Grassberger and I. Procaccia, "Characterization of attractors," *Phys. Rev. Lett.,* vol 50, No 346, pp. 346-349, 1983.

[8] P. Good, *Permutation, Parametric and Bootstrap Tests of Hypotheses*, 3rd ed., USA: Springer, 2004.

[9] A. Galka, *Topics in Nonlinear Time Series Analysis with Implications for EEG*. Singapore: World Scientific, 2000.

[10] D. Popivanov and A. Mineva, "Testing procedures for non stationarity and non-linearity in physiological signals," *Mathematical Biosciences,* vol 157, pp. 303-320, 1999.

[11] Y. Zhao, J. Sun and M. Small, "Evidence consistent with deterministic chaos in human cardiac data: Surrogate and nonlinear dynamical modeling," *International Journal of Bifurcation and Chaos,* vol 18, No 1, pp. 141-160, 2006.

[12] J. Barkoulas, "Testing for deterministic monetary chaos: Metric and topological diagnostics," *Chaos, Solitons and Fractals,* vol 38, pp. 1013-1024, 2008.

[13] M. Sano and Y. Sawada, "Measurement of the Lyapunov spectrum from chaotic time series," *Phys. Rev. Lett*, vol 55, No 10, pp. 1082-1085, 1985.

[14] C. Skokos, "The Lyapunov Characteristic Exponents and their Computation," Max Planck Institute for the Physics of Complex Systems, Germany, Report. 2009.

[15] R. Wayland, D. Bromley, D. Pickett and A. Passamante, "Recognozing determinism in a time series," *Phys. Rev. Lett.*, vol 70, pp. 580-582, 1993.

[16] A. R. Osborne and A. Provenzale, "Finite correlation dimension for stochastic systems with power law spectra," *Physica D*, vol 35, pp. 357-381, 1989.

[17] M. D. Nurujjaman, "Nonlinear dynamics experiments in glow discharge plasma," Ph.D. Thesis, Saha Institute of Nuclear Physics ,Calcula/India, 2009.

[18] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Phys. Rev. Lett*, vol 77, pp. 635-638, 1996.

[19] T. Theiler and D. Prichard, "Using "surrogate data" to calibrate the actual rate of false positives in tests for nonlinearity in time series," *Fields Inst. Comun*, 11, 99, 1997.

[20] E. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences,* vol 20, No 2, pp. 130-141, 1963.

[21] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, 1st ed., USA: Springer, 1996.

[22] D. Kugiumtzis, "Complications is applying the method of surrogate data to EEG," Max Planck Institute for the Physics of Complex Systems, Germany, Report, 1999.