

Análisis de dependencias no lineales utilizando redes neuronales artificiales

Analysis of nonlinear dependences using artificial neural networks

*Carlos A. Martínez, Juan D. Velásquez**

Facultad de Minas. Universidad Nacional de Colombia. A.A. 1027. Medellín, Colombia

(Recibido el 22 de septiembre de 2009. Aceptado el 01 de diciembre de 2010)

Resumen

En este artículo, se desarrolla una nueva técnica para detectar dependencias no lineales en series temporales, basadas en el uso de una red neuronal autorregresiva y el concepto de coeficiente de correlación. Teniendo en cuenta que el modelo de redes neuronales utilizado es capaz de aproximar cualquier función en un dominio compacto, las medidas propuestas son capaces de detectar no-linealidades en los datos. Nuestra técnica es probada para varios conjuntos de datos tanto simulados como reales, y comparada con las funciones clásicas de autocorrelación simple y parcial; los resultados muestran que en los casos lineales, las medidas propuestas tienen un comportamiento similar a las autocorrelaciones simple y parcial, pero en los casos no-lineales ellas son capaces de detectar otras relaciones no-lineales.

----- *Palabras clave:* red neuronal autorregresiva, modelado no lineal de series temporales, correlación múltiple, análisis de correlaciones en sistemas no lineales

Abstract

In this paper, we develop a new technique for detecting nonlinear dependences in time series, based on the use of an autoregressive neural network and the concept of coefficient of correlation. Taking into account that the employed neural network model is able to approximate any function in a compact domain, the proposed measures are able to detect nonlinearities in the data. Our technique is tested for various simulated and real datasets, and compared with classical functions of simple and partial autocorrelations; the results show that in the linear cases the proposed measures have a similar behaviour

* Autor de correspondencia: teléfono: +57 + 4 + 425 5371, fax: + 57 + 4 + 425 5365, correo electrónico: jdvelasq@unal.edu.co (J. D. Velásquez)

to the simple and partial autocorrelations, but in the nonlinear cases they are able to detect other nonlinear relationships.

----- *Keywords:* autoregressive neural network, nonlinear time series modelling, multiple correlation, analysis of correlation in nonlinear systems

Introducción

La caracterización, el modelado y la predicción de series temporales son tres problemas fundamentales en diversas áreas del conocimiento, tales como la ingeniería, la economía y las ciencias sociales [1]. La caracterización busca determinar las propiedades fundamentales de la serie bajo estudio [2]; en el modelado se pretende encontrar una descripción matemática (modelo) de la dinámica de largo plazo que sigue la serie, tal que se cumplan los supuestos básicos en que se basa el modelo usado. Finalmente, el objetivo de la predicción es desarrollar modelos que permitan determinar con la mayor precisión posible, los valores futuros de la serie investigada, para uno o más periodos hacia adelante.

Para atacar los problemas mencionados, se ha desarrollado numerosos modelos desde diferentes áreas como la Estadística y la Inteligencia Computacional, los cuales incluyen técnicas como los métodos estadísticos, las redes neuronales, los sistemas basados en reglas, y los sistemas híbridos [1].

Un paso fundamental dentro de prácticamente todas las metodologías de diseño de las técnicas mencionadas, corresponde a la determinación de las variables explicativas del modelo; véase por ejemplo [3, 4] entre muchos otros. Dichas variables explicativas pueden estar conformadas por los valores rezagados de la propia variable explicada o por los valores actuales o rezagados de otras variables exógenas al modelo. Box y Jenkins [3] sientan las bases fundamentales de la detección de dichas variables relevantes, al incorporar el uso de las funciones de autocorrelación simple y parcial, y correlaciones cruzadas simple y parcial, dentro de su metodología de especificación.

No obstante, el uso de las medidas de dependencia ya mencionadas es de utilidad dudosa cuando las relaciones existentes entre la variable dependiente y las variables explicativas son de carácter no-lineal. Particularmente, Granger y Lin [5] presentan ejemplos de relaciones determinísticas no-lineales para las cuales el autocorrelograma no presenta dependencias seriales significativamente diferentes de cero. Esta es la razón para que diversos investigadores hayan propuesto medidas alternativas a la correlación lineal para detectar relaciones no-lineales; por ejemplo, Cardona & Velásquez [6] analizan la selección de características relevantes en problemas de regresión usando información mutua; Granger y Teräsvirta [4] sugieren el uso del coeficiente de correlación calculado para distintas transformaciones tanto de la variable dependiente como de los regresores. Particularmente, Nielsen & Madsen [7] proponen el uso de una medida de correlación no lineal basada en el uso de una aproximación no paramétrica a partir de funciones de base entre la variable dependiente y el regresor considerado; la principal debilidad de esta propuesta está relacionada con la selección del número de funciones base a utilizar, y el cubrimiento sobre los datos que realiza cada una de ellas.

El objetivo principal de este trabajo es modificar la propuesta metodológica realizada por Nielsen & Madsen [7], reemplazando el aproximador propuesto por una red neuronal artificial autorregresiva (ARNN), la cual combina el modelo autorregresivo tradicional con un perceptrón multicapa [8 - 10]; y aplicar la propuesta realizada a la detección de la relación de dependencia para varios modelos. En la aproximación presentada en este artículo, el uso de la ARNN requiere que el analista defina únicamente la cantidad de neuronas en la capa oculta del modelo, por lo que evita entrar a

seleccionar el tipo de función base utilizada, la cantidad de funciones, y el cubrimiento de cada una de ellas, simplificando el proceso.

Para alcanzar los objetivos propuestos, el resto de este artículo está organizado como sigue. En la próxima sección, se discute el concepto de dependencia desde el punto de vista de los modelos de regresión, y se presentan algunas de las principales técnicas que se han propuesto para realizar su detección, haciendo énfasis en la propuesta metodológica de Nielsen & Madsen [7]. Posteriormente, se describe el modelo ARNN y se presenta la propuesta de aproximación metodológica. Después, se discuten los casos de aplicación desarrollados en este trabajo. Finalmente, se concluye en la última sección.

Técnicas de análisis de la dependencia entre dos variables

En esta sección se describen algunas de las principales técnicas que han sido propuestas en la literatura para la detección de la dependencia entre dos o más variables.

El uso de los correlogramas ha sido ampliamente difundido en la detección de relaciones lineales, debido principalmente, a que estas medidas de relación son una parte fundamental dentro de la estrategia de especificación de los modelos de Box y Jenkins [3]. Ellas permiten determinar que regresores tienen una influencia significativamente diferente de cero sobre la variable explicada, e igualmente detectar si existen correlaciones en los residuales del modelo. La función de autocorrelación simple permite medir el nivel de dependencia entre y_t y y_{t-k} cuando su relación es lineal.

El autocorrelograma simple corresponde al gráfico obtenido al relacionar el valor de la función de autocorrelación simple versus el rezago k . La función de autocorrelación parcial permite estimar numéricamente la relación entre y_t y y_{t-k} eliminando la influencia de los rezagos intermedios $y_{t-1}, \dots, y_{t-k-1}$. De manera similar, la función de correlación cruzada, c_k , mide la

dependencia lineal entre la variable dependiente y_t y una variable exógena x_{t-k} , mientras que la función de correlación cruzada parcial, realiza esta misma medición pero eliminando la influencia de otros regresores.

Granger y Teräsvirta [4] sugieren algunas medidas de correlación derivadas de las definiciones anteriores que podrían ser potencialmente útiles en el análisis de series temporales; todas ellas se basan en la transformación de la variable dependiente y el regresor considerado usando las funciones $g()$ y $f()$ respectivamente, tal que se maximiza la función de correlación. Específicamente, las funciones de correlación propuestas son las siguientes:

- El coeficiente de correlación máxima:

$$m\rho = \text{corr}(g(y), f(x)) \quad (1)$$

- El máximo de la correlación media:

$$mm = \text{corr}(y, f(x)) \quad (2)$$

- El máximo coeficiente de regresión:

$$mr = R^2 \quad (3)$$

- En la regresión de

$$y = f(x) + \text{residuo} \quad (4)$$

En donde las funciones $g()$ y $f()$ se escogen de tal forma que se maximice la medida de relación utilizada.

Tong [11] indica que el análisis de los gráficos de dispersión entre y_t y y_{t-k} es un método alternativo que puede ser útil para la detección de relaciones no lineales y comportamientos caóticos. Estos diagramas se construyen uniendo los puntos (y_t, y_{t-k}) y (y_{t+1}, y_{t+1-k}) mediante líneas.

Otra medida de la relación no lineal es la información mutua (IM); esta es una métrica que permite determinar la dependencia entre un conjunto de variables como una medida de distancia entre la distribución conjunta

actual entre los datos, $p(x, y)$ y la distribución que tendrían si ellos fuesen estadísticamente independientes, esto es, su distribución conjunta de probabilidades estaría dada por $p(x)p(y)$; esta medida de distancia es:

$$I(X;Y) = \int_x \int_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (5)$$

Cardona y Velásquez [6] analizan el uso de esta técnica en el caso de los modelos generales de regresión y desarrollan un estadístico para determinar cuando la IM es significativamente diferente de cero. La distribución de probabilidad de los datos se realiza mediante técnicas no paramétricas. La principal desventaja de esta medida de relación, está relacionada con la gran cantidad de datos requeridos para realizar una estimación suficientemente precisa de la distribución de probabilidades de los datos a través de técnicas no paramétricas.

Por otra parte, Nielsen y Madsen [7] indican que los coeficientes de correlación simple y parcial están estrechamente vinculados con el coeficiente de determinación. Así, el coeficiente de correlación múltiple al cuadrado, $\rho^2_{0(1\dots k)}$, entre la variable aleatoria dependiente Y y los regresores X_1, \dots, X_k puede ser calculado como:

$$\rho^2_{0(1\dots k)} = \frac{V[Y] - V[Y | X_1, \dots, X_k]}{V[Y]} \quad (6)$$

Donde $V[Y]$ es la varianza de la variable dependiente; $V[Y_t | Y_{t-k}]$ es la varianza de los residuales e_t .

El estimador de máxima verosimilitud de $\rho^2_{0(1\dots k)}$ calculado a partir de los datos es:

$$R^2_{0(1\dots k)} = \frac{S_0 - S_{0(1\dots k)}}{S_0} \quad (7)$$

Donde

$$SS_0 = \sum_{i=1}^N \left(y_i - \frac{\sum_{i=1}^N y_i}{N} \right)^2 \quad (8)$$

y $S_{0(1\dots k)}$ es la sumatoria de los residuos de la aproximación de y_i a partir de los regresores x_1, \dots, x_k . El coeficiente de correlación múltiple representa la reducción en la varianza de Y causada por los regresores x_1, \dots, x_k .

Igualmente, el coeficiente de correlación parcial, $\rho^2_{(0k)(1\dots k-1)}$, entre Y y X_k , dadas las variables X_1, \dots, X_{k-1} , permite medir la relación de dependencia entre Y y X_k una vez se ha eliminado el efecto de X_1, \dots, X_{k-1} . Esto es:

$$\rho^2_{(0k)(1\dots k-1)} = \frac{V[Y | X_1, \dots, X_{k-1}] - V[Y | X_1, \dots, X_k]}{V[Y | X_1, \dots, X_{k-1}]} \quad (9)$$

Su estimación a partir de la muestra de datos se realiza como:

$$R^2_{(0k)(1\dots k-1)} = \frac{S_{0(1\dots k-1)} - S_{0(1\dots k)}}{S_{0(1\dots k-1)}} \quad (10)$$

A partir de las definiciones anteriores, Nielsen y Madsen [7] definen la función de dependencia para el rezago k como:

$$LDF(k) = \text{sgn}(f_k(b) - f_k(a)) \sqrt{R^2_{0(1\dots k)}} \quad (11)$$

Donde a y b son los valores mínimo y máximo sobre las observaciones; $f()$ es un modelo de regresión que aproxima la relación entre Y y X_k . Igualmente, la función de dependencia parcial para el rezago k , se define como:

$$PLDF(k) = \text{sgn}(f_{kk}(b) - f_{kk}(a)) \sqrt{R^2_{(0k)(1\dots k-1)}} \quad (12)$$

Por definición $LDF(1) = PLDF(1)$.

Propuesta de aproximación metodológica

La aproximación metodológica presentada en este artículo, se basa en la utilización de una red neuronal autorregresiva para modelar la dependencia entre Y y X_k , que permite estimar las varianzas condicionales definidas en la ecuación (10). Así, $V[Y|X_1, \dots, X_k]$ puede ser estimada como la varianza de los residuales obtenidos al aproximar la variable Y mediante el modelo propuesto usando como regresores las variables X_1, \dots, X_k .

La red neuronal autorregresiva

Una red neuronal autorregresiva (ARNN) se obtiene al combinar un modelo lineal autoregresivo (AR) con un perceptron multicapa con una única capa oculta; véase [8-10]. Este es un modelo que permite combinar las ventajas de los modelos autoregresivos y de las redes neuronales, de tal forma que es más fácil capturar dinámicas complejas. Así para una ARNN, la variable dependiente y_t es obtenida después de aplicar la función no lineal:

$$y_t = \beta_* + \sum_{i=1}^I \varphi_i x_t^{(i)} + \sum_{h=1}^H \beta_h G \left((2\sigma_y)^{-1} \alpha_{*,h} + \sum_{i=1}^I \alpha_{i,h} x_t^{(i)} \right) + e_t \tag{13}$$

a un conjunto de regresores $x_t^{(i)}$ (para $i=1, \dots, I$), con $e_t = \sigma \varepsilon_t$, obtenida al asumir que los residuales e_t siguen una distribución normal con media cero y varianza desconocida. Los regresores pueden estar conformados por los retardos de y_t , así como por variables exógenas que pueden dar cuenta del comportamiento de y_t . $G()$ es la función de activación de las neuronas de la capa oculta. σ es la desviación estándar de los errores. ε_t es una variable aleatoria que sigue una distribución normal estándar. H es el número de neuronas en la capa oculta; I es el número de regresores. σ_y es la desviación estándar de y_t ; su uso evita tener que transformar y_t para restringir sus valores al rango de la función $G()$. Una representación pictórica de este modelo es presentada en la figura 1.

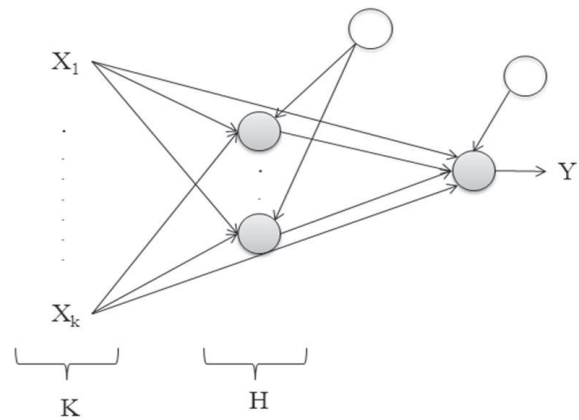


Figura 1 Red neuronal autorregresiva

Los parámetros del modelo ($\Omega = [\beta_*, \beta_h, \varphi_i, \alpha_{*,h}, \alpha_{i,h}]$, para $h=1, \dots, H$; $i=1, \dots, I$) son obtenidos maximizando el logaritmo de la función de verosimilitud de los errores:

$$L = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=1}^T \frac{e_t^2}{\sigma^2} \tag{14}$$

Mediante alguna técnica de optimización, usualmente basada en gradientes, teniendo en cuenta que T corresponde al número de observaciones. La maximización de (14) equivale a minimizar el error cuadrático medio, que es el procedimiento común en la literatura de redes neuronales.

Se propone la utilización de funciones tipo sigmoidea ya que algunos autores han sugerido por su experiencia práctica, que este tipo de funciones que son simétricas alrededor del origen convergen más rápidamente que la función sigmoidea tradicional; adicionalmente, la adición de un término lineal puede ayudar a la convergencia, ya que se evita la saturación de la neurona o unidad de procesamiento en la capa oculta, y garantiza un gradiente mínimo cuando la salida neta de la función sigmoidea es cercana a sus valores extremos. Consecuentemente con las razones expuestas, $G()$ es especificada como:

$$G(u) = \frac{1 - \exp(-u)}{1 + \exp(-u)} + \kappa u \tag{15}$$

Donde κ es una constante pequeña. Además de lo anterior, se sabe que para un conjunto cualquiera de parámetros Ω , los parámetros β_h pueden restringirse a ser positivos, ya que la contribución neta de cada unidad oculta, $\beta_h G()$, no cambia de signo si los parámetros $\beta_h, \alpha_{*h}, \alpha_{i,h}$ cambian de signo puesto que $\beta_h G(u) = -\beta_h G(-u)$. Igualmente, para evitar la multiplicidad de configuraciones en la capa oculta se puede obligar a que los parámetros β_h estén siempre ordenados de forma creciente: $0 < \beta_1 < \beta_2 < \dots < \beta_h$.

Al imponer que los parámetros φ_i sean igual cero, el modelo ARNN se reduce a un perceptron multicapa (MLP). Si H se hace igual a cero, (13) se reduce a un modelo autorregresivo lineal con entradas exógenas (ARX).

Estimación de las funciones de dependencia

En la propuesta de aproximación metodológica presentada, se definen las funciones de dependencia descritas a continuación a partir de las definiciones dadas en (6) y (9):

Función de autocorrelación no lineal entre y_t y y_{t-k} :

$$\rho_k = \frac{V[Y_t] - V[Y_t | Y_{t-k}]}{V[Y_t]} \quad (16)$$

Función de autocorrelación parcial no lineal entre y_t y y_{t-k} :

$$\rho_{kk} = \frac{V[Y_t | Y_{t-1}, \dots, Y_{t-k+1}] - V[Y_t | Y_1, \dots, Y_k]}{V[Y_t | Y_{t-1}, \dots, Y_{t-k+1}]} \quad (17)$$

Función de correlación cruzada simple no lineal entre y_t y x_{t-k} :

$$c_k = \frac{V[Y_t] - V[Y_t | X_{t-k}]}{V[Y_t]} \quad (18)$$

Función de correlación cruzada parcial no lineal entre y_t y x_{t-k} :

$$c_{kk} = \frac{V[Y_t | X_{t-1}, \dots, X_{t-k+1}] - V[Y_t | X_{t-1}, \dots, X_{t-k}]}{V[Y_t | X_{t-1}, \dots, X_{t-k+1}]} \quad (19)$$

Los valores numéricos de las funciones propuestas se calculan en forma similar a las definiciones dadas en (7) y (10). Así por ejemplo, en $16 V[Y_t]$ es la varianza de la variable dependiente; $V[Y_t | Y_{t-k}]$ es la varianza de los residuales e_t de un modelo ARNN que sigue la definición dada en (13), y para el cual $x_t = y_{t-k}$.

Estimación de los intervalos de confianza

Para obtener los intervalos de confianza se aplicó un contraste de permutación, el cual consiste en calcular el estadístico de interés (en este caso la función de correlación) sobre una muestra de datos en la cual la variable de salida es permutada. Al realizar este proceso N veces, se obtiene una muestra de tamaño N del estadístico en cuestión, la cual representa su distribución empírica de probabilidades. A partir de dicha muestra se pueden estimar los intervalos de confianza del estadístico estudiado. En todos los casos analizados $N=10000$. El proceso de cálculo es descrito en (6), entre muchos otros. Para agilizar el proceso de cálculo del intervalo de confianza, se puede asumir que el modelo ARNN no tiene neuronas en la capa oculta, por lo que se reduce a un modelo autorregresivo cuyos parámetros pueden ser calculados usando mínimos cuadrados.

Casos de aplicación

En esta sección se presentan los resultados obtenidos al aplicar las medidas de relación basadas en modelos de redes neuronales artificiales. En el primer caso se investigan las propiedades de la técnica propuesta, mientras que en los restantes se presentan casos reales de aplicación.

Caso 1. Investigación de las propiedades

En este primer caso, se desea investigar las propiedades numéricas de las medidas de autocorre-

lación no-lineal propuestas sobre una muestra de datos cuyo proceso generador es conocido. Los procesos generadores considerados que se presentan a continuación fueron usados por Granger y Lin [5] en su estudio sobre la Información Mutua. Los modelos son los siguientes:

- (M1) $y_t = e_t + 0.8 e_{t-3}^2$ (NLMA(3))
- (M2) $y_t = 0.6 e_{t-1} y_{t-2} + e_t$ (bilineal)
- (M3) $y_t = 4y_{t-1}(1 - y_{t-1})$ (caos determinístico)
- (M4) $y_t = |y_{t-1}|^{0.8} + e_t$ [NLAR(1)]
- (M5) $y_t = \text{sign}(y_{t-1}) + e_t$ [NLAR(1)]
- (M6) $y_t = y_{t-1} + e_t$ [paseo aleatorio]

Para cada uno de los modelos considerados se estimaron 200 muestras, cada una con un tamaño

de 300 datos. Para cada una de las series obtenidas se estimaron las funciones ya indicadas. En la tabla 1 se presentan los valores promedios de las funciones de autocorrelación simple lineal (ACSL) y no-lineal (ACSNL) estimados sobre las 200 muestras respectivamente. e_t es un variable aleatoria normal estándar. Tal como se puede observar en la tabla 1, las funciones propuestas permiten detectar las dependencias no lineales existentes en los datos; esto es especialmente importante para el modelo M3, en que la ACSL no indica la presencia de ninguna relación de dependencia, mientras que la ACSNL indica que hay una dependencia muy importante para el rezago 1. Para el proceso de paseo aleatorio (M6) las funciones de ACSL y ACSNL presentan un comportamiento decreciente muy similar, tal como se espera para este tipo de procesos.

Tabla 1 Valores promedios de las funciones de autocorrelación no lineal para las funciones del Caso 1

Rezago	M1	M2	M3	M4	M5	M6
	ACSL	ACSL	ACSL	ACSL	ACSL	ACSL
	(ACSNL)	(ACSNL)	(ACSNL)	(ACSNL)	(ACSNL)	(ACSNL)
1	-0,027 (0,027)	-0,016 (-0,045)	-0,073 (-0,972)	0,450 (0,604)	0,543 (0,678)	0,927 (0,935)
2	0,004 (-0,040)	-0,052 (-0,080)	-0,027 (-0,033)	0,189 (0,320)	0,322 (0,364)	0,868 (0,870)
3	-0,063 (-0,116)	0,022 (-0,048)	0,068 (-0,120)	0,076 (0,179)	0,178 (0,207)	0,811 (0,808)
4	-0,013 (-0,041)	0,004 (0,017)	-0,046 (0,000)	0,049 (0,037)	0,079 (0,065)	0,756 (0,751)
5	-0,004 (0,016)	0,019 (0,073)	-0,023 (-0,030)	0,005 (-0,013)	0,039 (0,045)	0,704 (0,691)
6	-0,012 (-0,013)	-0,043 (0,015)	0,031 (-0,141)	-0,030 (0,037)	0,028 (0,007)	0,659 (0,642)

ACSL: autocorrelación simple lineal; ACSNL: autocorrelación simple no lineal

Caso 2. Procesos lineales y no lineales

En este segundo caso de aplicación se estiman los autocorrelogramas lineales y no-lineales para las

funciones analizadas en [7] para una muestra de 100 datos. Los procesos considerados en [7] son los siguientes:

(M7) $Y_t = 1.13Y_{t-1} - 0.64Y_{t-2} + e_t$ (AR(2))

(M8) $Y_t = e_t + 0.698e_{t-1} - 0.5247e_{t-2}$ (MA(2))

(M9) $Y_t = \frac{1}{1 + \exp(-5Y_{t-1} + 2.5)} + e_t$
(autorregresivo no lineal)

(M10) $Y_t = e_t + 2 \cos(e_{t-1})$
(promedio móvil no lineal)

En la figura 2, se presenta la realización de cada proceso que fue considerada. Los autocorrelogramas calculados para cada uno de los procesos son presentados en la figura 3. Para el proceso M7, los autocorrelogramas simples muestran claramente la influencia de los dos primeros rezagos de y_t ; mientras que los autocorrelogramas parciales difieren principalmente en el signo de la relación (rezago 2).

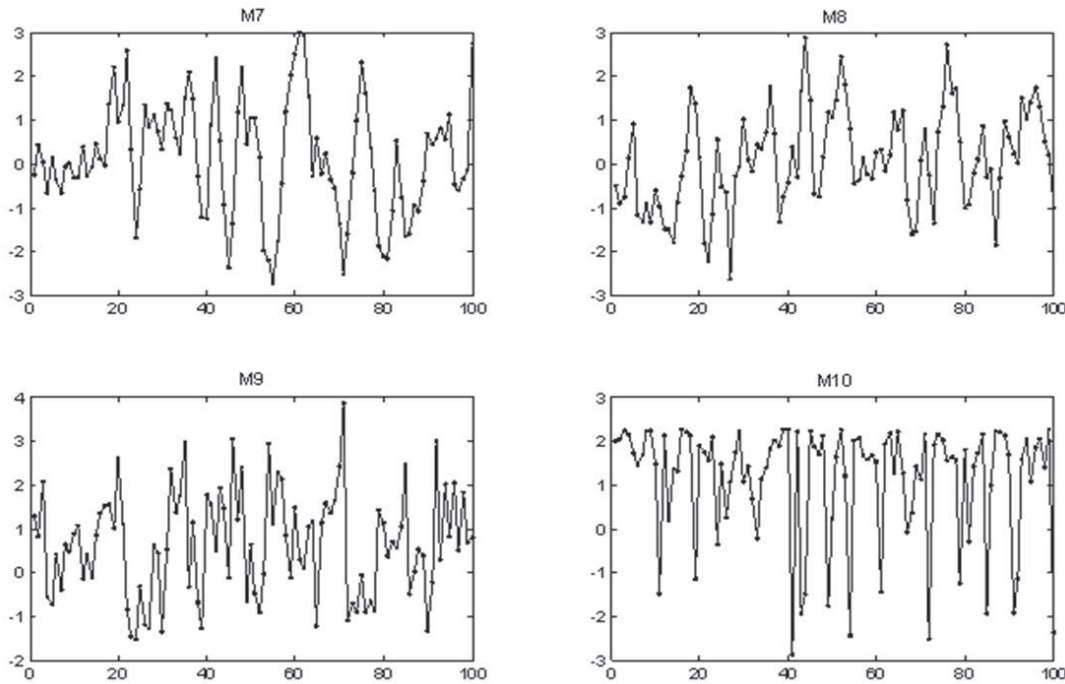


Figura 2 Realizaciones de los procesos analizados en el Caso 2

Caso 3. The canadian lynx data

La serie corresponde al número de Lynx atrapados por año en el distrito del río Mackenzie en Canadá. La serie comprende 114 observaciones correspondientes al periodo comprendido entre 1821 y 1934, y ha sido comúnmente utilizada para realizar comparaciones entre modelos de predicción no lineales [12-15]. Los datos usados comúnmente para el modelado corresponde a la serie original transformada; su gráfica aparece en la figura 4. La serie exhibe una componente cíclica de periodo igual a 10 años,

la cual es irregular en amplitud. En la figura 5 se presentan los autocorrelogramas simples y parciales, lineales y no lineales, calculados para la serie transformada; en este caso se usó un modelo ARNN con 2 neuronas en la capa oculta para todos los casos considerados. Las autocorrelaciones simples presentan un comportamiento notoriamente similar, y muestran claramente la componente cíclica de periodo 10. La función de autocorrelación no lineal muestra valores más bajos sobre la parte inferior del autocorrelograma en relación al

autocorrelograma lineal. Esto puede ser causado en parte, a que el número de neuronas utilizado en la construcción de la medida no lineal puede ser insuficiente para capturar el comportamiento de la serie; Zhang [14] ha reportado resultados para un perceptrón multicapa con una estructura de 7 entradas, 5 neuronas en la capa oculta y una

salida. La función de autocorrelación parcial lineal indica que los rezagos 1, 2, y 11 son relevantes, mientras que podría existir una influencia no tan importante por parte de los rezagos 3, 4 y 7. En contraste, la función de autocorrelación parcial no lineal indica que existe una clara influencia de los rezagos 1, 2, 3, 4, 7, 8, 12, 13.

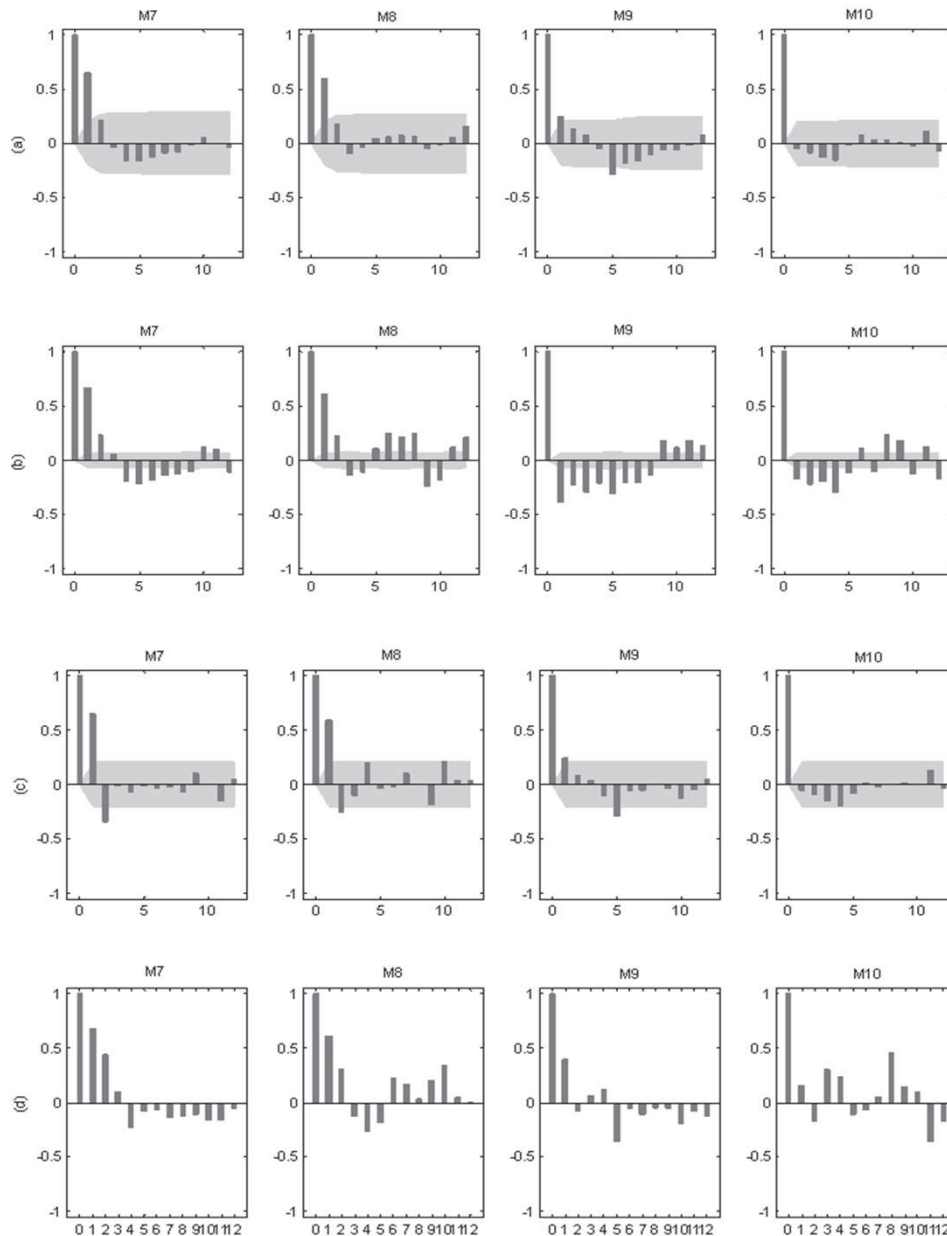


Figura 3 Funciones de relación no lineal para el Caso 2. (a) Autocorrelograma simple lineal; (b) Autocorrelograma simple no lineal; (c) Autocorrelograma parcial lineal; y (d) Autocorrelograma parcial no lineal

Caso 4. Número de manchas solares de Wolfer

Esta serie corresponde al número de grupos de manchas solares visibles sobre la superficie del sol por año, registradas entre 1700 y 1987; véase la figura 6. Los estudios realizados por diferentes autores indican que la serie es no lineal y que la distribución de probabilidades de sus datos no es gaussiana, por lo que ha sido utilizada para la comparación entre modelos no lineales [14, 16, 17]. Las configuraciones típicas presentadas en la literatura usan los siguientes grupos de rezagos de la variable dependiente como entradas al modelo:

- 1-4
- 1, 2, 9, 11
- 1-11
- 1-3, 9-11

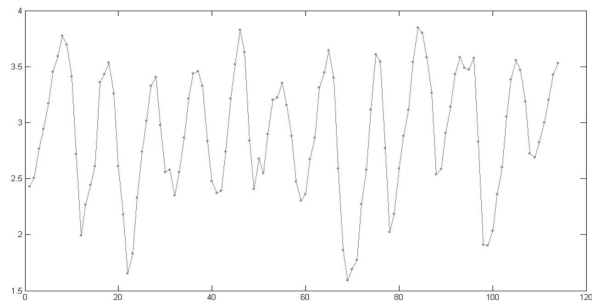


Figura 4 Serie Canadian Lynx transformada

En la figura 7 se presentan los autocorrelogramas estimados para la serie. Los autocorrelogramas simples muestran un comportamiento bastante similar; igualmente, los autocorrelogramas parciales muestran un comportamiento similar, excepto por la diferencia en el signo para las medidas en algunos rezagos.

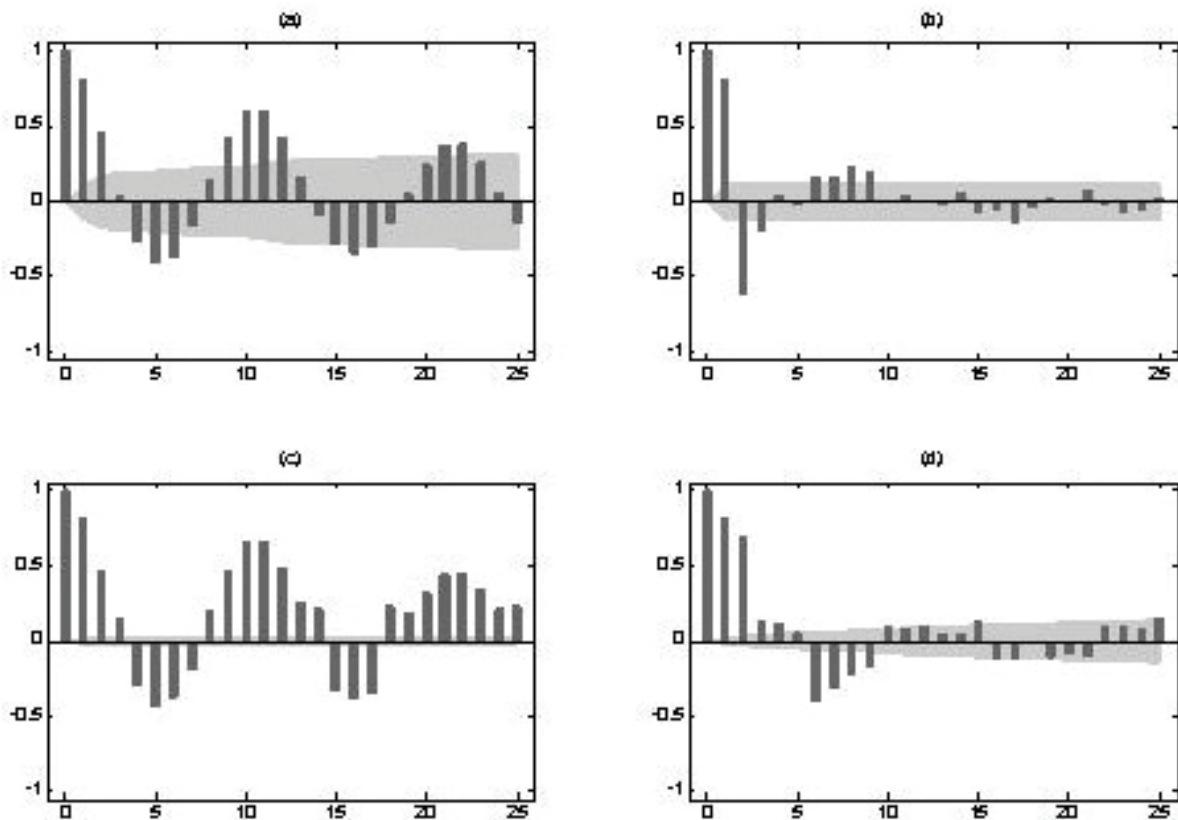


Figura 5 Autocorrelogramas para la serie de manchas solares anuales. (a) Autocorrelograma simple lineal; (b) Autocorrelograma simple no lineal; (c) Autocorrelograma parcial lineal; y (d) Autocorrelograma parcial no lineal

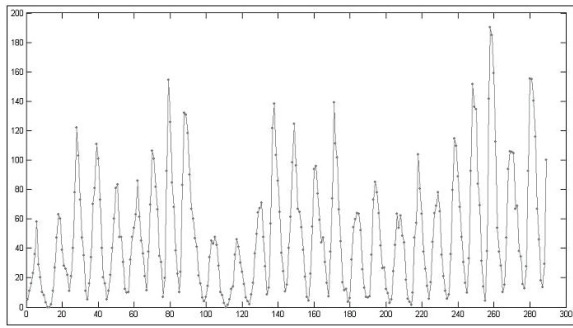


Figura 6 Manchas solares anuales

Conclusiones

En este artículo, se desarrolla una metodología para la detección de relaciones lineales y no lineales en series temporales, a partir de la definición del concepto de coeficiente de correlación, y el uso de una red neuronal autorregresiva. El modelo propuesto permite superar las falencias

encontradas en las medidas basadas en correlación lineal, y puede ser usado para la determinar que variables deberían ser incorporadas formalmente dentro del proceso de especificación de un modelo no-lineal. Su uso no está limitado a los ejemplos ilustrados en este trabajo, y la técnica propuesta podría ser utilizada para determinar otro tipo de relaciones no lineales tales como la heterocedasticidad en los residuales. Las medidas definidas son dependientes de la capacidad que tengan los modelos de redes neuronales para aproximar la relación existente, de tal forma, que la calidad de los resultados es dependiente de la calidad del algoritmo utilizado para la estimación de los parámetros del modelo ARNN; consecuentemente, se hace necesario investigar cuales algoritmos podrían ser los más adecuados para entrenar la red neuronal artificial tanto en términos de tiempo como de calidad del punto de mínima encontrado a partir de la función de error definida.

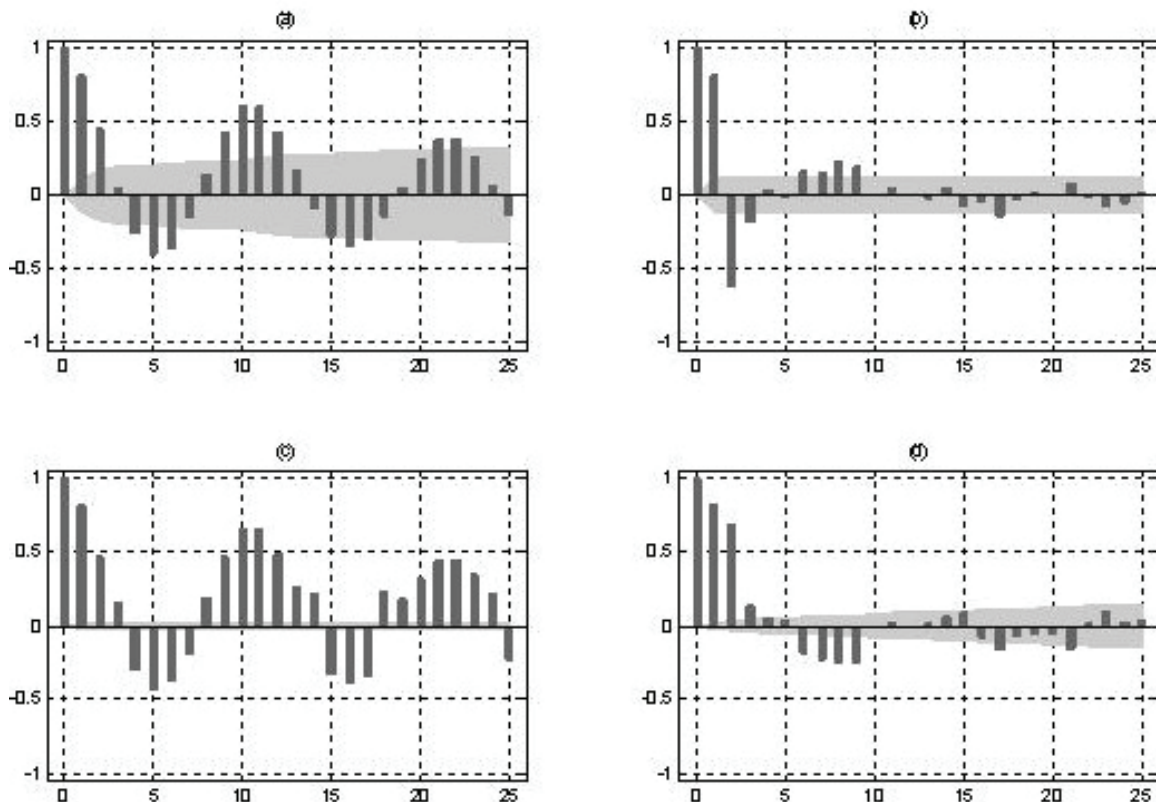


Figura 7 Autocorrelogramas para la serie de manchas solares anuales. (a) Autocorrelograma simple lineal; (b) Autocorrelograma simple no lineal; (c) Autocorrelograma parcial lineal; y (d) Autocorrelograma parcial no lineal

Referencias

1. N. Kasabov. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. 2nd ed. Ed. Massachusetts Institute of Technology. 1998. pp. 43-47.
2. A. Weigend, N. Gersfeld. *Time-Series Prediction: Forecasting the future and understanding the past*. Reading, Mass. Ed. Addison-Wesley. 1993. pp. 5-35.
3. G. E. P. Box, G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Ed. Holden-Day Inc. 1970. pp.1-17.
4. C. Granger, T. Teräsvirta. *Modeling Nonlinear Economic Relationships*. Ed. Oxford University Press. 1993. pp.25-35.
5. C. Granger, J. Lin. "Using the mutual information coefficient to identify lags in nonlinear models". *Time Ser. Anal.* Vol. 15. 1994. pp. 371-384.
6. C. A. Cardona, J. D. Velásquez. "Selección de características relevantes usando información mutua." *Revista Dyna*. Vol.73. 2006. pp. 149-163.
7. H.A. Nielsen, H. Madsen. "A generalization of some classical time series tools". *Computational Statistics & Data Analysis*. Vol. 37. 2000. pp.13-31.
8. H. White. *An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks*. In Proceedings of the International Joint Conference on Neural Networks. Washington, DC, IEEE Press, NY. 1989. Vol. 2. pp. 451-455.
9. T. H. Lee, H. White, C. W. J. Granger. "Testing for neglected nonlinearity in time series models." *Journal of Econometrics*. Vol. 56. 1993. pp. 269-290.
10. T. Teräsvirta, C. F. Lin, C. W. J. Granger. "Power of the neural network linearity test." *Journal of Time Series Analysis*. Vol. 14. 1993. pp. 209-220.
11. H. Tong. *Non-linear Time Series, a dynamical system approach*, Oxford Statistical Science Series, Ed. Claredon Press Oxford. 1990. pp. 215-238.
12. M. J. Campbell, A. M. Walker. "A survey of statistical work on the mackenzie river series of annual Canadian lynx trappings for the years 1821-1934 and a new analysis." *Journal of the Royal Statistical Society: Series A. Statistics in Society*. Vol. 140. 1977. pp. 411-431.
13. T. S. Rao, M. M. Gabr. "An introduction to bispectral analysis and bilinear time series models." *Lecture Notes in Statistics*. Vol. 24. 1984. pp. 528-535.
14. G. Zhang. "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing*. Vol. 50. 2003. pp. 159-175.
15. M. Ghiassi, H. Saidane. "A dynamic architecture for artificial neural network". *Neurocomputing*. Vol. 63. 2005. pp. 397-413.
16. M. Cottrell, M. Girard, Y. Girard, M. Mangeas, C. Muller. "Neural modeling for time series: a statistical stepwise method for weight elimination." *IEEE Transactions on Neural Networks*. Vol. 6. 1995. pp. 1355-1364.
17. C. de Groot, D. Wurtz. "Analysis of univariate time series with connectionists nets: a case study of two classical examples." *Neurocomputing*. Vol. 3. 1991. pp. 177-192.