

A comparison of time series forecasting between artificial neural networks and box and jenkins methods

Joanna Collantes Duarte¹, Gerardo Colmenares La Cruz², Giampaolo Orlandoni Merli³ y Franklin Rivas Echeverría⁴

¹Escuela de Estadística. Facultad de Ciencias Económicas y Sociales.

²Instituto de Investigaciones Económicas. ³IEAC-CESIMO. ⁴Laboratorio de Sistemas Inteligentes. Universidad de los Andes. Mérida, Venezuela. E-mail: joannac@ula.ve; orlandon@ula.ve; gcolmen@ula.ve; rivas@ula.ve

Abstract

This paper deals with a comparison between Box and Jenkins methodologies and Artificial Neural Networks on time series forecasting. ARIMA and Transfer Function Models are compared with Neural Network Models. Performance of the building models are analysed using comparative criteria during the prediction and fitness stage.

Key words: Forecasting time series, Box and Jenkins methodology, ARIMA model, transfer function model, artificial neural network, neo-fuzzy neuron.

Predicción con redes neuronales artificiales: comparación con las metodologías de box y jenkins

Resumen

El objetivo principal de esta investigación es comparar las metodologías de Box y Jenkins: Modelo ARIMA y Modelo de Función de Transferencia (MFT), utilizadas frecuentemente en estadística para predicción con series de tiempo, con la técnica de la Inteligencia Artificial denominada Redes Neuronales Artificiales (RNA). Se proponen metodologías para predicción con una red neuronal artificial utilizando el algoritmo de retropropagación y las neuronas neo-difusas. El comportamiento de los métodos se analiza mediante casos de estudio y haciendo uso de criterios comparativos para las fases de ajuste y predicción.

Palabras clave: Predicción con series de tiempo, metodología de Box y Jenkins, modelo ARIMA, modelo de función de transferencia, redes neuronales artificiales, neurona neo-difusa.

1. Introducción

La predicción basada en series de tiempo es de gran interés práctico, pues permite conocer, con un margen de error, valores futuros de una serie basándose en sus valores pasados (caso univariable), o valores pasados de una serie ayudan a predecir valores futuros de otra serie (caso bivariable).

La estadística clásica utiliza las metodologías de Box y Jenkins (BJ) [1] para resolver pro-

blemas de predicción. La aplicación de estas metodologías puede resultar compleja, lo cual conduce a intentar dar solución a este problema mediante las nuevas técnicas computacionales que han tomado auge en los últimos años.

La Inteligencia Artificial ha surgido como una nueva área del conocimiento. Está formada por un conjunto de técnicas que intentan imitar, en forma artificial, las habilidades relacionadas con la inteligencia humana. Una de estas técni-

cas es la denominada Redes Neuronales Artificiales (RNA), utilizadas con un relativo éxito para la predicción de series de tiempo. Pueden citarse algunos trabajos en el área como los realizados por: Wong [2], Hill et al [3], Wedding II y Cios [4], Faraway y Chatfield [5], entre otros.

El objetivo fundamental de este trabajo es comparar, mediante casos de estudio, los métodos de predicción: RNA (con algoritmo de aprendizaje: retropropagación y neurona neodifusa) con las metodologías de BJ: modelo ARIMA (caso univariable) y MFT (caso bivariable), utilizando como criterios comparativos los errores de ajuste y predicción.

El artículo está organizado como sigue: La sección 2 presenta las metodologías de BJ: el modelo ARIMA y el modelo de función de transferencia. La tercera sección es una introducción a las redes neuronales, específicamente a las redes unicapas que utilizan el algoritmo de retropropagación y se introducen las características principales de las neuronas neodifusas. Se propone, en la sección cuatro, una metodología para predicción, para el caso univariable y bivariable. La quinta sección corresponde a los experimentos y resultados; se aplica la metodología ARIMA y RNA a la serie de número de nacimientos mensuales y el modelo de función de transferencia y RNA a las series de gastos mensuales en publicidad y número de ventas mensuales. Se analizan estos dos casos mediante varios criterios comparativos, tanto en la fase de ajuste como de predicción. Por último, en la sección seis se establecen algunas conclusiones.

2. Series de Tiempo: Metodologías de Box y Jenkins

Una serie de tiempo discreta es una secuencia cronológica de observaciones de una variable particular [6]. El objetivo, más frecuente, en series de tiempo es predecir valores futuros de una serie a partir de sus valores pasados. Estos valores futuros pueden obtenerse intentando conseguir un patrón en los datos y haciendo extrapolaciones para conseguir tales predicciones. El método univariable de BJ es ampliamente utilizado en la estadística clásica para la predicción con series de tiempo de una variable [1], basado en el proceso Autorregresivo Integrado de Promedio Móvil (ARIMA: **A**utoregressive **I**ntegrated **M**oving **A**verage). Variantes de esta metodología conside-

ran los modelos de función de transferencia, usados para predecir valores de una serie de tiempo, a partir de valores pasados de esa serie y otras series con relación causal. En esta sección se estudiarán estas dos metodologías.

2.1. Modelo ARIMA

El Método ARIMA de Box y Jenkins (BJ) requiere que la serie sea estacionaria, esto significa que su media, varianza y covarianza permanezcan constantes sin importar el momento en el cual se midan. Para convertir una serie no estacionaria en estacionaria se propone el proceso de diferenciación [7]. Esta metodología aplica modelos autorregresivos (AR), de promedio móvil (MA) y modelos mixtos, tanto a la parte estacional de la serie (se le antepone una S al modelo para indicar que corresponde a la parte estacional) como a la parte no estacional. Los modelos estacionales consideran los retrasos del proceso y la perturbación aleatoria periódicamente, es decir, cada "S" periodos.

La unión de modelos estacionales con modelos no estacionales conduce a un modelo de gran capacidad de adaptación que puede reflejar tanto la tendencia como la estacionalidad de una serie. Estos modelos se generan mediante la multiplicación de los operadores polinomiales que caracterizan a cada modelo, obteniéndose los modelos conocidos como $ARIMA(p,d,q) \times (P,D,Q)_S$, donde los parámetros no considerados en el modelo seleccionado serán de orden cero. Este modelo puede expresarse como:

$$\phi_p(B^S)\phi_p(B)(1-B)^d(1-B^S)^D y_t = \delta + \theta_q(B^S)\theta_q(B)a_t \quad (1)$$

donde:

- $\phi_p(B)$ $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ Modelo AR
- $\theta_q(B)$ $(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ Modelo MA
- $\phi_p(B^S)$ $(1 - \phi_1 B^S - \phi_2 B^{2S} - \dots - \phi_p B^{pS})$ Modelo SAR
- $\theta_q(B^S)$ $(1 - \theta_1 B^S - \theta_2 B^{2S} - \dots - \theta_q B^{qS})$ Modelo SMA
- y_t Valores originales.
- B Operador retardo ($BY_t = Y_{t-1}$).
- S Periodo estacional.
- p Orden del modelo AR.
- q Orden del modelo MA.
- d Orden de la diferenciación en la parte no estacional.
- P Orden del modelo SAR.

- Q Orden del modelo SMA.
- D Orden de la diferenciación en la parte estacional.
- a_t Perturbación aleatoria. Ruido Blanco.
- δ Constante

El orden de la parte AR y MA del modelo final no son escogidos arbitrariamente. Para esto BJ proporcionan un método estructurado que permite determinar cuál es el modelo que mejor se ajusta a la serie en cuestión y se recomienda que el modelo se mantenga tan simple como sea posible (principio de parsimonia). En general, no hay más de tres términos AR o MA [4]. Las herramientas principales para determinar el orden de cada modelo son los correlogramas simple y parcial, que son la representación gráfica de las funciones de autocorrelación simple y parcial [1].

Etapas de la Metodología ARIMA

La construcción de un modelo ARIMA puede ser estructurada en cinco etapas:

1. Análisis Exploratorio de la Serie: Consiste en un primer análisis, mediante gráficos y pruebas estadísticas para convertir la serie en estacionaria (media y varianza constante).
2. Identificación del Modelo: Se definen los conjuntos de estimación y predicción. Haciendo uso de los correlogramas simple y parcial, se identifican posibles modelos en la parte no estacional y estacional.
3. Estimación de los Parámetros del Modelo: Los valores de los parámetros ϕ y θ se estiman mediante la minimización de la suma de cuadrados de los errores ϵ_t , mediante el método de mínimos cuadrados condicional, no condicional o máxima verosimilitud [1, 13, 15].
4. Adecuación del Modelo: Existen varias características resaltantes: a) Se estudian los residuos, obtenidos por la diferencia entre el valor original de la serie y el valor estimado por el modelo, deben aproximarse al comportamiento de un ruido blanco (media cero, varianza σ^2 y covarianza cero). b) Los parámetros del modelo ARIMA seleccionado, deben ser significativamente diferentes de cero y estar poco relacionados entre sí. c) El grado de ajuste de este modelo debe ser elevado en comparación al de otros modelos

alternativos. d) La bondad del ajuste puede evaluarse con la Desviación Estándar Residual (DER), Criterio de Información de Akaike (AIC) y con el Criterio Bayesiano de Schwarz (SBC), entre otros [5].

5. Predicción: Mediante el modelo ARIMA seleccionado se predicen "m" períodos, correspondientes al tamaño del conjunto de predicción, con sus intervalos de confianza y se calculan los correspondientes errores de predicción. Es importante juzgar la adecuación del modelo en función de qué tan bien se pronostican los datos no empleados para la estimación del modelo. Para evaluar la capacidad de predicción, se calcula el error medio absoluto (EMA) y la raíz del error cuadrático medio porcentual (RECOMP) [8].

2.2. Modelo de Función de Transferencia

Con el Modelo de Función de Transferencia (MFT) en la metodología de BJ se predicen valores futuros de una serie de tiempo a partir de ella misma, y de otras relacionadas. Es decir se considera el caso bivariable donde un modelo puede predecir valores futuros de una serie de tiempo (serie de salida), a partir de valores pasados de esa serie y otra serie de tiempo relacionada (series de entrada) [6]. Se parte del supuesto que existe una relación causal unidireccional *a priori*, desde la serie de entrada hacia la serie de salida, eliminando la posibilidad de retroalimentación.

Etapas para construir un modelo de función de transferencia

La metodología para construir Modelos de Función de Transferencia (MFT), basada en BJ y propuesta por Bowerman y O'Connell, es la siguiente:

1. Identificación de un modelo que describa a la serie de entrada y preblanqueo de la serie de entrada y salida.
Se denota al t -ésimo término de la serie de entrada con x_t , y al de la serie de salida con y_t . Se asume que la transformación necesaria para hacer a la serie de entrada estacionaria es la misma que para la serie de salida. Sea la serie de entrada estacionaria, $Z_t^{(x)}$. Se deberán seguir los siguientes pasos:

- a) Análisis de los correlogramas simple y parcial de la serie de entrada x_t , para determinar la necesidad de alguna transformación que conduzca a que la serie de tiempo sea estacionaria $Z_t^{(x)}$.
- b) Análisis de los correlogramas simple y parcial de $Z_t^{(x)}$ para identificar el modelo ARIMA (p, q, P, Q).
- c) Estimación del modelo. Se selecciona un modelo cuyos parámetros sean significativos.

$$\phi_p(B^s) \phi_p(B) Z_t^{(x)} = \theta_q(B^s) \theta_q(B) \varepsilon_t \quad (2)$$

- d) Adecuación del modelo mediante el análisis de los residuos (igual al caso univariable).
- e) Preblanqueo de x_t ($Z_t^{(x)}$). Se despeja ε_t de la ecuación (2). A este error del modelo para la serie de entrada se le denomina α_t .

$$\alpha_t = \frac{\hat{\phi}_p(B^s) \hat{\phi}_p(B)}{\hat{\theta}_q(B^s) \hat{\theta}_q(B)} Z_t^{(x)} \quad (3)$$

- f) Preblanqueo de y_t ($Z_t^{(y)}$). Al error del modelo para la serie de salida se le denomina β_t .

$$\beta_t = \frac{\hat{\phi}_p(B^s) \hat{\phi}_p(B)}{\hat{\theta}_q(B^s) \hat{\theta}_q(B)} Z_t^{(y)} \quad (4)$$

- 2. Identificación de un MFT preliminar que describa la serie de entrada.

Se calcula la función de correlación cruzada (CC), $r_k(\alpha_t, \beta_t)$, entre los valores de α_t y los valores de β_t en el retraso k; esta CC es una medida de la relación lineal entre los valores de α_t y β_{t+k} y se representa mediante el gráfico de CC. En base a la función de CC se procede a identificar los parámetros del MFT (b,s,r):

- a) Verificar que no existan valores significativos en las CC antes del retraso cero (en los MFT se asume que valores presentes de y_t están relacionados con valores presentes y/o pasados de x_t y no al contrario).
- b) Identificar b, que es el retraso donde se observa el primer valor de r_k estadísticamente diferente de cero.
- c) Identificar s, que es el número de valores pasados de x que influyen sobre y. El valor de s es igual al número de retrasos que están en-

tre el primer r_k significativo y el comienzo del patrón de decaimiento.

- d) Identificar r, que representa el número de valores pasados de y que están relacionados con y_t . Se determina examinando el decaimiento de la CC después del retraso (b+s). Si cae en forma exponencial amortiguada, entonces r será igual a 1 y si cae en forma sinusoidal amortiguada, el valor de r será 2.
- e) El MFT preliminar general es de la forma:

$$Z_t^{(y)} = \mu + \frac{Cw(B)}{\delta(B)} B^b Z_t^{(x)} + \eta_t \quad (5)$$

donde:

- μ Constante que se incluye en el modelo si es diferente de cero
- C Parámetro escalar desconocido
- η_t Variable aleatoria independiente de X
- w(B) $1 - w_1B - w_2B^2 - \dots - w_sB^s$
- $\delta(B)$ $1 - \delta_1B - \delta_2B^2 - \dots - \delta_rB^r$

- 3. Identificación de un modelo que describa a η_t y modelo final de función de transferencia.

- a) Calcular la función de CC de los residuos con los valores de entrada x_t y las autocorrelaciones simple y parcial de los residuos.
- b) Verificar que se cumpla una condición necesaria para la validez del MFT, en la que se establece que la serie de entrada preblanqueada α_t debe ser estadísticamente independiente del componente del error η_t . Esto se cumple cuando los valores de la probabilidad en las correlaciones cruzadas son grandes y no se puede rechazar la hipótesis de independencia.
- c) Encontrar, por medio de los correlogramas simple y parcial de los residuos, el modelo que describe a η_t .

$$\eta_t = \frac{\hat{\theta}_q(B^s) \hat{\theta}_q(B)}{\hat{\phi}_p(B^s) \hat{\phi}_p(B)} a_t \quad (6)$$

donde a_t es una perturbación aleatoria en el tiempo t, con distribución ruido blanco.

- d) Se sustituye η_t en el MFT preliminar (ecuación 5) constituyendo el modelo final de función de transferencia:

$$Z_t^{(y)} = \mu \frac{Cw(B)}{\delta(B)} B^b Z_t^{(x)} + \frac{\hat{\theta}_q(B^s) \hat{\theta}_q(B)}{\hat{\phi}_p(B^s) \hat{\phi}_p(B)} a_t \quad (7)$$

Una vez construido el MFT se generan los valores correspondientes al ajuste y predicción de la serie y_t , y se calculan sus respectivos errores, como en la metodología ARIMA.

3. Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA), inspiradas en las neuronas biológicas, persiguen imitar ciertas habilidades humanas atribuibles al cerebro y a millones de elementos interconectados llamados Neuronas. Las neuronas son células nerviosas que constituyen los elementos primordiales del sistema nervioso central. Éstas son capaces de recibir señales, procesarlas y transmitir las a otras neuronas [9, 10].

Una neurona artificial es una unidad de procesamiento de información que es fundamental para la operación de una red neuronal [10]. En un modelo esquemático de la neurona pueden identificarse esencialmente seis elementos. Las entradas: escalares que se le proporcionan a la red, de acuerdo al problema en estudio; las salidas: valores que determina la red como resultado de su aprendizaje; los pesos sinápticos: arreglo de valores numéricos que expresa la importancia del enlace neuronal correspondiente entre la entrada y la salida (el valor de la entrada x_i es asociado con la neurona "k" mediante el peso sináptico w_{ik}); punto de suma de entradas ponderadas

(función de acumulación): combinación lineal o suma de todas las entradas multiplicadas por sus correspondientes pesos; función de activación: función que puede ser lineal o no lineal y limita el rango de la salida de la neurona; y sesgo: señal de entrada adicional que consiste en un valor fijo denominado x_0 (generalmente igual a 1) y un peso adicional w_{0k} .

En la figura 1 se ilustra un modelo de neurona, donde se observa que la neurona k puede describirse mediante las funciones de acumulación y activación:

$$v_k = \sum_{i=0}^p x_i w_{ik} \quad \text{y} \quad y_k = \Gamma(v_k) \quad (8) \text{ y } (9)$$

donde x_0, x_1, \dots, x_p son las entradas de la neurona k; $w_{0k}, w_{1k}, \dots, w_{pk}$ son los pesos de la neurona k; v_k es la suma de las entradas multiplicadas por los pesos correspondientes; $\Gamma(\cdot)$ es la función de activación y y_k es la salida de la neurona k. Los pesos son parámetros escalares que se van ajustando según la aplicación de una regla de aprendizaje, de manera de cumplir con la relación entrada/salida, y la función de activación se selecciona de acuerdo al objetivo del problema y al requerimiento de la salida. La función logística, como función de activación, se recomienda para problemas de predicción. Esta función toma valores entre 0 y 1 y su expresión matemática es:

$$\Gamma(v) = \frac{1}{1 + e^{-v}} \quad (10)$$

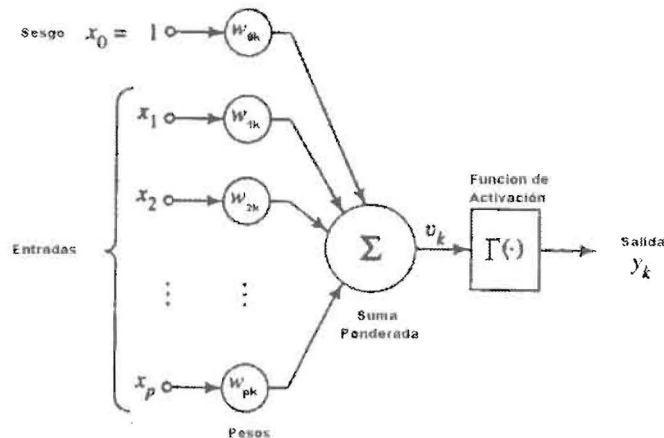


Figura 1. Modelo de una Neurona.

La construcción de RNA puede incluir capas ocultas (no están directamente conectadas a entradas o salidas), donde las neuronas en cada capa tienen la misma estructura de una sola neurona, pero sus entradas son las salidas de las neuronas de la capa anterior. La regla de aprendizaje o algoritmo de entrenamiento es un procedimiento para modificar los pesos de una red y su propósito es entrenar la red para ejecutar alguna tarea [11]. El algoritmo de retropropagación es el más usado para entrenar RNA multicapas de alimentación adelantada.

El método general de entrenamiento puede resumirse en los siguientes pasos:

Pasos hacia delante (desde la entrada hacia la salida):

1. Seleccionar un patrón de entrada del conjunto de entrenamiento.
2. Aplicar esta entrada a la red y calcular la salida.

Pasos hacia atrás (desde la salida hacia la entrada):

3. Calcular el error entre la salida de la red neuronal y la salida deseada para el patrón de entrada usado.

4. Ajustar los pesos para que el error cometido entre la salida de la red neuronal y la salida deseada sea disminuido.

5. Repetir los pasos 1 al 4 para todos los patrones de entrenamiento, hasta que el error global sea aceptablemente bajo.

El criterio de parada se aplica cuando los valores de los pesos reducen la función del error (diferencia cuadrática entre la salida deseada y la salida de la red neuronal), y este proceso se realiza por el método del gradiente descendente. La idea del método es realizar un cambio en los pesos proporcional a la derivada de la función de error respecto al peso para cada patrón.

Durante la fase de asociación, la red opera enteramente en forma de cascada directa o hacia adelante, es decir sin retroalimentación. Sin embargo, el ajuste de los pesos obtenido por la regla de entrenamiento se realiza desde atrás hacia delante, pasando por las capas ocultas hasta el nivel de entrada [9], por lo cual se le denominó a este procedimiento "algoritmo de retropropagación".

Una vez que una red neuronal ha sido entrenada, deberá estar en capacidad de proveer salidas próximas a los valores deseados cuando se le proporcionan nuevos ejemplos, es decir, entradas que no pertenecen al conjunto de entrenamiento, sino que forman parte de los patrones de prueba. A este proceso se le conoce como generalización de un modelo de RNA.

Neuronas Neo-Difusas

La Figura 2 ilustra la estructura de una neurona Neo-Difusa, donde los pesos de interconexión (sinapsis) han sido sustituidos por un conjunto de funciones no lineales f_i , y el cuerpo celular, como en la mayoría de los esquemas neuronales, realizan la suma de las señales sinápticas [18].

La Figura 3 ilustra la estructura de las funciones no lineales f_i . Estas funciones están com-

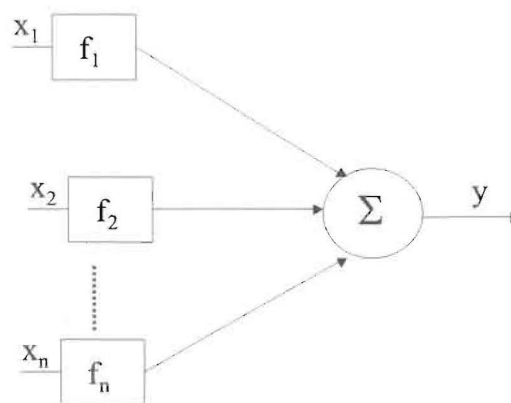


Figura 2. Estructura de una Neurona Neo-Difusa.

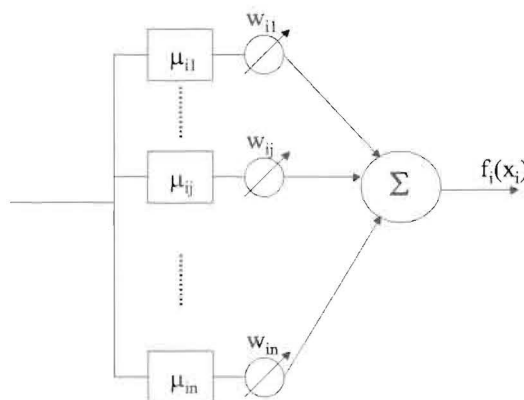


Figura 3. Estructura de la función no lineal (Sinapsis).

puestas de reglas del tipo: SÍ <condición> ENTONCES <acción>; utilizando <condición> como el grado de pertenencia que posee cada una de las señales de entrada que están incluidas en cada una de los *segmentos difusos complementarios* construidos de la forma representada en la Figura 4. La <acción> es el correspondiente valor de w_{ij} .

El valor de salida ($f_i(x_i)$) de la sinapsis no lineal es obtenida por un proceso de desfuzificación que considera la estructura complementaria de los segmentos (esto significa que la suma de las dos funciones de pertenencia activadas es igual a 1). Así, la salida de la neurona Neo-Difusa está dada por:

$$f_i(x_i) = \mu_{ik}(x_i)w_{ik} + \mu_{i,k+1}(x_i)w_{i,k+1}$$

donde

$\mu_{ik}(x_i)$ es el valor de pertenencia generado por la señal de entrada x_i .

w_{ik} son los pesos de interconexión.

El algoritmo de entrenamiento de *actualización incremental* [18] utilizado para el ajuste de pesos, es de la siguiente manera:

$$\Delta w_{ij} = -\alpha(y_k - t_k)\mu_{ij}(x_{ik})$$

donde:

y_k es la salida de la Neurona Neo-Difusa.

t_k es la salida deseada

α es la tasa de aprendizaje

4. Una Metodología para Predicción con Redes Neuronales Artificiales

A continuación se presenta una metodología para construir, entrenar y probar una red neuronal artificial para predicción, tanto para el caso de una sola serie de tiempo (modelo univariable), como para el caso de una serie de tiempo que tiene una relación causal con otra (modelo bivariado).

1. Escalamiento de los datos: Existen diversas técnicas para el escalamiento de los datos, tales como la estandarización (la mayoría de valores escalados se encuentran entre -3 y 3), escalamiento entre -1 y 1, escalamien-

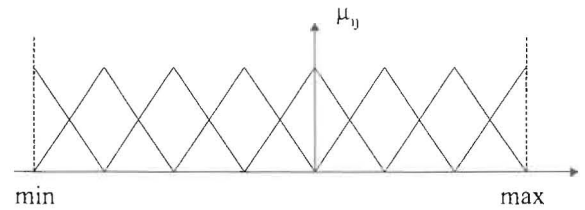


Figura 4. segmentos difusos complementarios.

to entre 0 y 1, entre otros. La idea de estas técnicas es llevar a todas las entradas de la RNA a órdenes de magnitud similares, para evitar que exista una predisposición numérica a ponderar de mayor forma aquellas variables con valores numéricos más grandes. Una de las técnicas más utilizadas consiste en transformar los datos de la(s) serie(s) a valores comprendidos entre 0 y 1, mediante:

$$y_{i(\text{escalados})} = \frac{y_i - \text{Min}}{\text{Max} - \text{Min}} \quad (11)$$

2. Patrones de Entrenamiento y Prueba: Los valores de la(s) serie(s) de tiempo se dividen en dos conjuntos de datos: a) Patrones de entrenamiento: conjunto formado por el 80% de los datos de la(s) serie(s), que se seleccionan en forma consecutiva y ordenada. Este conjunto de datos se utilizará para el entrenamiento de la red neuronal. b) Patrones de prueba: conjunto formado por el 20% de los datos de la(s) serie(s), correspondiente a los datos restantes, una vez que se han seleccionado los patrones de entrenamiento. Este conjunto de datos se utiliza para evaluar la capacidad de generalización o predicción de la red.
3. Topología de la RNA: Dirección de la información: Alimentación adelantada. Tipo de interconexión: Totalmente conectada. N° de entradas: $p + 1$ (una constante de valor 1, denominada sesgo o intercepto). N° de capas ocultas: 1. N° de nodos en la capa oculta: q . N° de salidas: 1. Función de activación de los nodos de la capa oculta: logística. Función de activación de la salida: logística.

En el caso de las redes neo-difusas se usa una sola neurona y el único parámetro ajustable es el número de segmentos.

4. Determinación de las p entradas a la red neuronal artificial:

Caso Univariable:

- La periodicidad de los datos: se consideran 12 o 13 retrasos debido a que las series son mensuales. Pudiera ser conveniente construir una primera red con 13 entradas, correspondientes a los 13 retrasos y analizar los pesos, de manera que se seleccionen las entradas asociadas a los pesos de mayor magnitud, como lo sugieren Faraway y Chatfield [5]. Este análisis de los pesos ayuda a identificar las variables de entrada más importantes.
- Una vez determinado el modelo ARIMA, seleccionar como entradas los valores correspondientes a los retrasos de y_i involucrados en este modelo. Esto se realiza con la intención de verificar hasta que punto un conocimiento estadístico *a priori*, puede ayudar en el diseño de una RNA para predicción de series de tiempo.
- Otra alternativa es la de considerar los correlogramas simple y parcial de la serie estacionaria. Las entradas a la red serían los datos correspondientes a los retrasos que resultaran con una correlación significativamente diferente de cero.
- Mediante pruebas por ensayo y error: considerando, por ejemplo, los datos retrasados 1,4,8,12 periodos, 1,3,6,9,12 periodos, o 1,2,3,4 periodos.

Caso Bivariable:

Se deben considerar retrasos de la serie a predecir (y), así como también valores presentes y pasados de la serie relacionada (x). Para definir las entradas pueden considerarse los siguientes aspectos:

- Periodicidad de los datos. Para datos mensuales se consideran 12 o 13 retrasos por serie.
- Los retrasos de interés para las series de entrada y salida pueden identificarse, como en el caso univariable, mediante el modelo ARIMA o mediante los correlogramas simple y parcial (se seleccionan los retrasos cuyas correlaciones sean significativamente diferentes de cero).

- Modelo de Función de Transferencia: se considera como entrada los valores de la serie de entrada (x) correspondientes a los b retrasos, indicados en el MFT. También los valores de la serie de salida (y) correspondientes a los r retrasos.
 - Correlación Cruzada: se consideran como entradas los valores correspondientes a los retrasos de la serie x que tengan CC significativamente diferentes de cero.
 - Mediante pruebas por ensayo y error, considerando diferentes retrasos para ambas series.
5. Determinación del número de nodos de la capa oculta (q). Una regla *ad hoc*, que en experimentos previos ha resultado de utilidad, asume que el valor inicial del número de nodos de la capa oculta sea igual al promedio entre el número de entradas y salidas. Se puede complementar la regla mediante pruebas por ensayo y error, agregando más nodos, y comparando los errores de ajuste y predicción. Esto no aplica en el caso de las neuronas neo-difusas ya que están formadas por una sola neurona.
6. Algoritmo de Entrenamiento: Retropropagación [9, 10, 11]. El algoritmo de entrenamiento para las neuronas neo-difusas es el algoritmo de actualización incremental [18].
7. Selección de los pesos iniciales. Es recomendable probar con diferentes conjuntos de valores iniciales para tratar de obtener buenos resultados. Los pesos iniciales se generan aleatoriamente en cincuenta oportunidades [5]. Se selecciona el modelo que obtenga el menor promedio entre la suma de cuadrados de los errores de ajuste y predicción.
8. Entrenamiento de la RNA seleccionada. Se establece el número máximo de ciclos, el error permitido de convergencia, tasa de aprendizaje, el incremento de las tasas de aprendizaje y momento se fija por ensayo y error. Una vez definido el modelo, se generan los valores de la serie de tiempo ajustada o producida por la red, utilizando los patrones de entrenamiento. Finalmente, se calcula el error de entrenamiento.
9. Predicción. Usando el mejor modelo previamente entrenado, se obtiene el valor de pre-

dicción y_{t+1} . Para hacer predicciones más allá de este período, se utiliza y_{t+1} como entrada para producir la predicción y_{t+2} y así sucesivamente para todo el conjunto de predicción. Al finalizar se calcula el error de predicción, como medida de generalización del modelo.

10. Comparación entre modelos con diferentes entradas. Usando el error de entrenamiento y de generalización de cada modelo se hace la comparación y se selecciona el modelo de red cuyos valores sean mínimos. La correlación entre los valores originales de la serie y los estimados puede usarse como una medida de la exactitud de la predicción. De este modo se selecciona el modelo cuyos errores de entrenamiento y predicción sean menores.

A continuación se muestra un ejemplo de una RNA para predicción (caso univariable) con una capa oculta, Figura 5.

5. Experimentos y Resultados

Se estudiarán dos aplicaciones, una correspondiente al caso univariable y la otra al bivariable. Para el análisis de los resultados obtenidos en los diferentes métodos se consideran los siguientes criterios comparativos:

- a) Criterios comparativos para el ajuste [5]:
- Desviación Estándar Residual (DER): Es un criterio de selección de modelo y un valor pequeño indica una mayor adecuación del modelo. Su expresión es:

$$DER = \sqrt{\frac{S}{T-r}} \quad (12)$$

donde S es la suma cuadrática de los residuos, T es el número de las observaciones efectivas que se usan en el ajuste y r es el número de parámetros estimados en el modelo, incluyendo la constante.

- Criterio de Información de Akaike:

$$AIC = T \ln(S/T) + 2r \quad (13)$$

Este criterio permite seleccionar un modelo. Se prefiere el modelo que tenga el menor valor de AIC.

- Criterio Bayesiano de Schwarz (SBC):

$$SBC = T \ln(S/T) + r \ln(T) \quad (14)$$

Este criterio es un método para la selección de un modelo. Se prefiere el modelo que tenga el mínimo valor del estadístico. El SBC penaliza los parámetros adicionales más severamente que el AIC, conduciendo a modelos más simples.

- b) Criterios comparativos para la predicción:

- Error Medio Absoluto (EMA).

$$EMA = \frac{1}{m} \sum_{i=1}^m |y_{t+i} - \hat{y}_{t+i}| \quad (15)$$

donde y_{t+i} son los valores observados de la serie que pertenecen al conjunto de predicción y, \hat{y}_{t+i} son los valores pronosticados por el modelo ARIMA [8].

- Raíz del Error Cuadrado Medio Porcentual (RECMP) [8]:

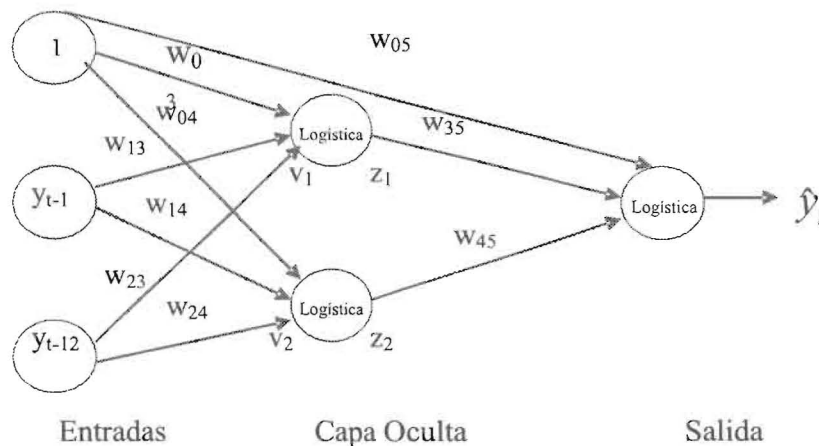


Figura 5. RNA de alimentación adelantada con una capa oculta para predicción de series de tiempo.

$$\sqrt{\frac{1}{m} \sum_{i=1}^m \left(\frac{y_{t+i} - \hat{y}_{t+i}}{y_{t+i}} \right)^2} \quad (16)$$

Correlación (R) entre los valores observados y los pronosticados por el modelo. Valores altos de esta correlación indican una buena adecuación del modelo.

$$R = \frac{\text{Covarianza}(y, \hat{y})}{\sqrt{\text{Varianza}(y) * \text{Varianza}(\hat{y})}} \quad (17)$$

**5.1. Caso Univariable:
Serie Nacimientos**

En el caso de predicciones con una sola serie de tiempo se analizará la serie de tiempo del número de nacimientos mensuales, ocurridos en España desde enero de 1960 hasta diciembre de 1999, lo que equivale a 40 años [12]. Estos datos se dividen en dos conjuntos: datos para el ajuste o entrenamiento: enero 1960 - diciembre 1991, comprende 384 meses (32 años) que equivale al 80% de los datos y datos para la predicción o prueba: enero 1992 - diciembre 1999, comprende 96 meses (8 años) que equivale al 20% de los datos. La serie completa se presenta gráficamente a continuación: Figura 6.

El análisis se hará aplicando las metodologías ARIMA y RNA. Al aplicar la metodología ARIMA (sección 2.1), utilizando el paquete SPSS [13, 14], se obtienen los correlogramas: Figura 7 y 8.

En la parte regular se observa un patrón MA (Promedio Móvil) de orden 1 (q=1). En la parte estacional se observa patrón autoregresivo y de promedio móvil. Sin embargo, el modelo de mejor adecuación y predicción resultó ser el ARIMA(0,1,1)*(0,1,1)₁₂ cuyo modelo estimado es [15]:

$$(1-B^{12})(1-B)y_t^* = (1-0,712 B^{12})(1-0,744 B)\epsilon_t \quad (18)$$

donde y_t^* es la serie transformada por logaritmo natural.

Para el análisis de los residuos se utilizó la prueba de Kolmogorov Smirnov que permite verificar si los residuos siguen una distribución Normal. Se obtuvo una significación de 0,999, lo que indica que no puede rechazarse la hipótesis nula de normalidad de los residuos. La media de los errores es -0,00045, aproximadamente cero. En los correlogramas de los residuos se observan po-

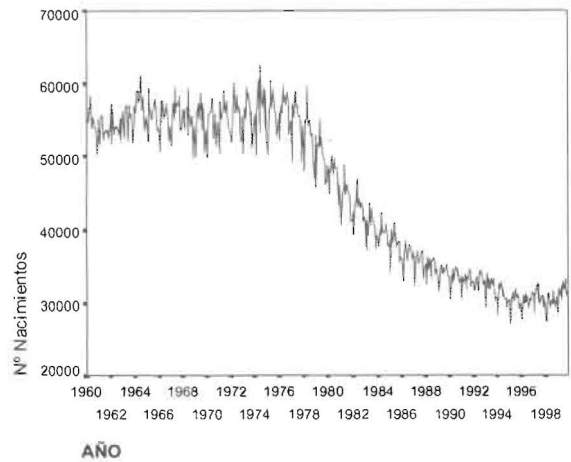


Figura 6. Grafico de la Serie de Tiempo del Número de Nacimientos Mensuales, ocurridos en España desde Enero de 1960 hasta Diciembre de 1999.

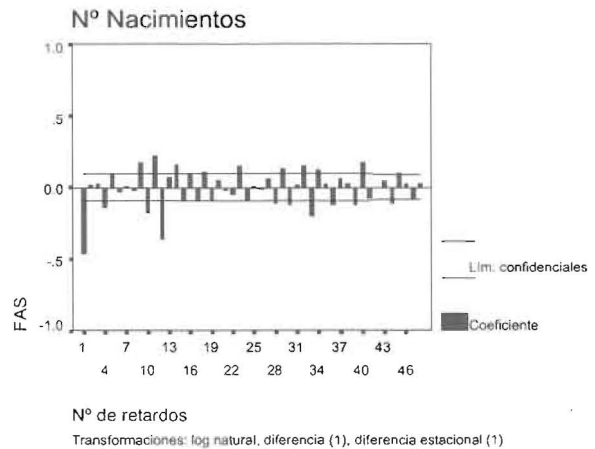


Figura 7. Correlograma Simple con las transformaciones: logaritmo natural, 1 era. diferencia en la parte regular y 1 era. diferencia en la parte estacional.

cos valores significativos como se muestra a continuación: Figura 9 y 10.

Se aplica la metodología de RNA para el caso univariable (sección 4), mediante el paquete Matlab [16]. Una vez realizadas las pruebas por ensayo y error para el ajuste de parámetros, se seleccionaron 5 modelos, cuyas entradas se especifican en la Tabla 1.

Al comparar mediante las medidas de ajuste y predicción los modelos de RNA y los de Redes Neo-Difusas (RND), se obtuvo: Tabla 2.

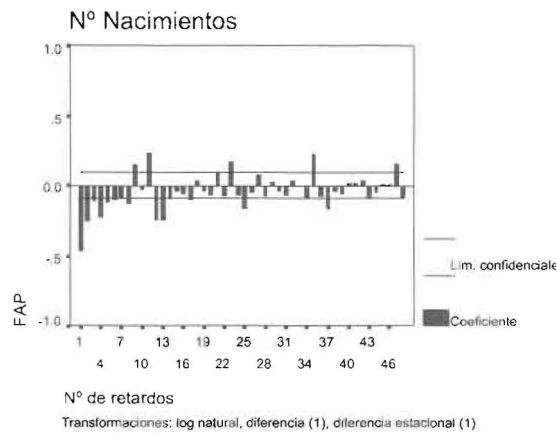


Figura 8. Correlograma Parcial con las transformaciones: logaritmo natural, 1 era. diferencia en la parte regular y 1 era. diferencia en la parte estacional.

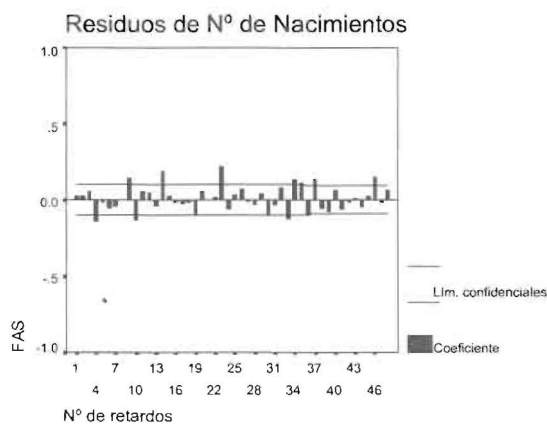


Figura 9. Correlograma Simple de los residuos para la serie nacimientos.

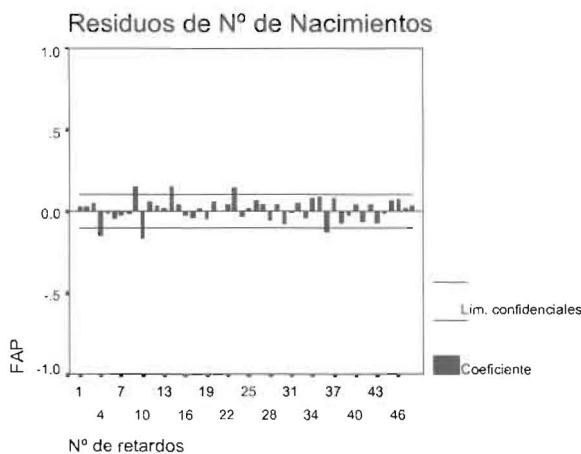


Figura 10. Correlograma Parcial de los residuos para la serie nacimientos.

De los modelos de RNA se selecciona el modelo RNA6 por obtener las mejores medidas de ajuste y predicción. En relación a los modelos de RND se selecciona el modelo RND6 por alcanzar la mejor predicción.

En la Tabla 3 se presentan las medidas de ajuste (entrenamiento) y predicción (prueba), tanto para el Modelo ARIMA seleccionado como para los mejores modelos de RNA y RND obtenidos.

El modelo con mejor predicción es el RNA6 que tiene como entradas los tres primeros retrasos, mientras que el modelo de mejor ajuste es el modelo ARIMA. Los errores de ajuste son menores en el modelo ARIMA, lo que implica un mejor ajuste en la fase de entrenamiento, mientras que las medidas de predicción son mejores en el modelo RNA6, indicando una mejor predicción o generalización con el uso de la RNA. Obsérvese que tanto en el modelo de RNA como en el RND se obtienen mejores predicciones que en el modelo ARIMA, lo cual es una mejora importante porque los modelos de RNA6 y RND6 utilizan solamente data reciente (los 3 primeros retrasos) de la serie de tiempo, lo que implica una más fácil implantación y más patrones de entrenamiento pueden obtenerse de la data.

Comparando los modelos RNA y RND, puede observarse que la mejor predicción fue alcanzada por el modelo de RNA, ratificando el buen desempeño de la RNA en tareas de predicción, pero el modelo de RND arrojó un resultado bastante aceptable, mejor que el obtenido con el modelo ARIMA, usando una estructura tan simple como una sola neurona.

5.2. Caso Bivariable: Serie Publicidad y Ventas

Para el caso bivariable se analizaron los datos presentados por Makridakis, Whelwright y McGee referenciados por Bowerman y O'Connell [6]. Estos datos consisten de 100 observaciones sobre dos series de tiempo: número de ventas mensuales (en miles de casos) y gasto mensual en publicidad (en miles de dólares). El objetivo es predecir el número de ventas (y_t) en base a sus valores pasados y a la publicidad (x_t), considerando que existe una relación causal entre ambas. Para construir el modelo de función de transferencia (MFT) se siguió la metodología se-

Tabla 1
Entradas de los modelos de RNA (caso univariable)

Modelo	Entradas (retrasos)	Justificación
RNA1	Del 1 al 13	Periodicidad de los datos
RNA2	1, 12 y 13	Retrasos del modelo ARIMA
RNA3	1, 2, 4 y del 9 al 13	Correlaciones significativas en los correlogramas simple y parcial
RNA4	1, 2, 12 y 13	Se agregó el retraso 2 a los utilizados en el modelo ARIMA
RNA5	3, 5 y 9	Según el análisis de los pesos de la RNA que considera los retrasos del 1 al 13: como los pesos entre la capa oculta y la de salida son bastante homogéneos, se considera la matriz de pesos entre las entradas y la capa oculta, se calcularon las medias de los pesos en cada entrada y se seleccionaron las tres entradas con medias mayores en valor absoluto.
RNA6	1,2 y 3	Los 3 valores anteriores

ñalada en la sección 2.2, haciendo uso del SPSS [13] y SAS [17]. La serie de entrada x_t corresponde al gasto mensual en publicidad y la serie de salida y_t al número de ventas mensuales. El MFT final obtenido fue el siguiente [15]:

$$z_t^{(y)} = \frac{1,974}{1 - 0,460 B} B^2 z_t^{(x)} + \frac{1 - 0,534 B}{1 - 0,534 B^3} a_t \quad (19)$$

Para la construcción de los modelos de RNA se consideraron los retrasos especificados en la Tabla 4.

Se compararon los seis modelos de RNA seleccionados a través de los errores de ajuste y predicción (Tabla 5).

Se seleccionó el modelo RNA2 por obtener la mejor predicción (menor EMA y RECOMP, y la más alta correlación entre los valores verdaderos y la predicción del modelo RNA entrenado). Podría considerarse el modelo RNA4, pues tiene el menor AIC y los demás valores están cerca de los encontrados para el modelo RNA2. En este modelo RNA2 se consideraron como entradas los retrasos que resultaron con correlaciones significativamente diferentes de cero en los correlogramas simple y parcial del análisis univariable de las series. El modelo RNA4 tiene como entradas los retrasos de la serie publicidad que resultaron significativamente diferentes de cero en las correlaciones cruzadas entre ambas series, así como los retrasos correspondientes a las correlaciones

significativamente diferentes de cero en los correlogramas simple y parcial de la serie ventas.

En Tabla 6 se presentan los errores de ajuste durante el entrenamiento y predicción, como medida de generalización, tanto para el MFT seleccionado como para el mejor modelo de RNA obtenido.

Se observa que el modelo RNA2 tiene mejor desempeño que el obtenido con el MFT. Inclusive, en general, si se observa la tabla 5, el desempeño del modelo RNA5 (las entradas son los retrasos correspondientes al MFT) es mejor, cuando se compara con el MFT.

6. Conclusiones y Recomendaciones

Una de las limitaciones de las RNA para alcanzar un uso generalizado es el establecimiento de la topología de la red. En este trabajo se proporciona una metodología que pudieran servir de guía en la predicción con series de tiempo. El punto más álgido en el diseño del modelo de RNA está en la selección de las entradas o retrasos de la serie. En forma empírica, se llegó a la conclusión de que puede utilizarse el método ARIMA como una herramienta de preprocesamiento de datos, considerando como entradas los retrasos involucrados en este modelo, lo cual sugiere que el uso de los conocimientos de métodos estadísticos, aplicados en el diseño de RNA, puede ayudar

Tabla 2
Errores de ajuste y predicción para modelos de RNA y de RND.

RNA Nº	Entradas (Retrasos)	Nº Nodos Capa Oculta	Ajuste			Predicción		
			DER	AIC	SBC	EMA	RECOMP	R
RNA1	1-13	7	2.977	6.147	6.562	1.083	4,47	0,54
RNA2	1,12,13	2	2.365	5.786	5.829	996	4,07	0,66
RNA3	1,2,4,9-13	4	2.949	6.010	6.170	1.133	4,82	0,50
RNA4	1,2,12,13	2	2.920	5.946	5.997	1.113	4,64	0,53
RNA5	3,5,9	2	2.639	5.867	5.910	985	4,19	0,65
RNA6	1,2,3	2	1.703	5.543	5.586	863	3,49	0,70
RND1 (2 segmentos)	1,12,13	1	2.458	5.813	5.852	1.469	5,70	0,44
RND2 (4 segmentos)	1,12,13	1	2.445	5.819	5.878	1.128	4,74	0,45
RND3 (6 segmentos)	1,12,13	1	2.581	5.865	5.936	1.148	4,69	0,45
RND4 (2 segmentos)	1,2,3	1	1.993	5.657	5.697	1.358	5,21	0,46
RND5 (4 segmentos)	1,2,3	1	2.040	5.685	5.743	1.110	4,59	0,51
RND6 (6 segmentos)	1,2,3	1	2.155	5.731	5.802	1.104	4,45	0,51

Tabla 3
Medidas de ajuste y predicción para los mejores modelos RNA, RND y ARIMA

RNA Nº	Ajuste			Predicción		
	DER	AIC	SBC	EMA	RECOMP	CORR
RNA6	1.703	5.543	5.586	863	3,49	0,70
RND6	2.155	5.731	5.802	1.104	4,45	0,51
ARIMA	1.372	5.362	5.370	1.173	5,09	0,48

a proponer soluciones mejor estructuradas. En el caso del MFT no se puede llegar a la misma conclusión; seguramente, porque en el ejemplo estudiado no se logra captar el patrón de comportamiento del proceso. Sin embargo, resultaron ser muy útiles los correlogramas simple y parcial, y la correlación cruzada, para identificar los retrasos a ser entradas.

En el ejemplo estudiado para el caso univariable, el modelo ARIMA proporcionó un mejor

ajuste, mientras que la RNA con un ajuste similar, pero cuantitativamente inferior, logra mejores predicciones. El modelo ARIMA no logra captar el cambio de patrón que ocurre al final de la serie, pudiendo ser corregida mediante un Análisis de Intervención [1]. En contraste, las RNA si logran captar este cambio. Una razón para justificar este fenómeno podría ser la no linealidad del modelo de RNA, y la linealidad del modelo ARIMA. El modelo ARIMA encontrado requiere los retra-

Tabla 4
Entradas de los modelos de RNA (caso bivariable)

Modelo	Entradas Retrasos serie xt	Entradas Retrasos serie yt	Justificación
RNA1	Del 1 al 13	Del 1 al 13	Periodicidad de los datos
RNA2	1, 2, 3, 5 y 6	1 y 2	Correlaciones significativas en los correlogramas simple y parcial de cada serie.
RNA3	1	1 y 2	Retrasos de los modelos ARIMA encontrados
RNA4	2, 3 y 6	1 y 2	Correlaciones cruzadas significativas
RNA5	2	1	Considerados por el MFT
RNA6	1, 2 y 3	1, 2 y 3	Prueba por ensayo y error

Tabla 5
Errores de ajuste y predicción para los seis modelos de RNA seleccionados

RNA N°	Entradas		N° de Nodos Capa Oculta	Ajuste			Predicción		
	Publicidad	Ventas		DER	AIC	SBC	EMA	RECOMP	R
RNA1	1- 13	1 - 13	13	3,9	911,7	1.716,4	10,3	5,8	0,97
RNA2	1,2,3,5,6	1,2	4	6,1	315,7	397,3	6,3	3,8	0,98
RNA3	1	1,2	2	10,1	332,2	356,4	10,6	6,0	0,96
RNA4	2,3,6	1	2	5,8	280,5	329,0	8,2	4,2	0,98
RNA5	2	1	1	15,4	376,1	387,2	21,3	10,6	0,87
RNA6	1,2,3	1,2,3	3	5,6	280,3	335,4	8,7	4,9	0,97

Tabla 6
Errores de ajuste y predicción para los mejores modelos de RNA y MFT

RNA N°	Ajuste			Predicción		
	DER	AIC	SBC	EMA	RECOMP	CORR
RNA2	6,1	315,7	397,3	6,3	3,8	0,98
MFT	18,26	463,25	544,83	39,35	19,18	-0,22

los 1, 12 y 13 para crear el modelo de predicción, esta característica reduce el tamaño del conjunto de estimación, mientras que los otros dos modelos utilizan los retrasos 1, 2 y 3. El modelo de RND alcanza una capacidad de predicción mejor que la lograda por el modelo ARIMA, lo que puede

considerarse un logro para una estructura tan simple como la de este modelo: una sola neurona.

En el ejemplo utilizado para el caso bivariable, las RNA resultaron tener mejor desempeño que el MFT, tanto en el ajuste como en la predicción. El hecho de dividir el conjunto de datos en

80% para el ajuste y el 20% restante reservarlo para predicción afecta al MFT, que cambia radicalmente con relación al modelo obtenido por Bowerman y O'Connell [6] utilizando el 100% de los datos para el ajuste. Esto podría ser un indicativo de la sensibilidad del MFT a la cantidad de información suministrada, y de la robustez de la RNA para captar la estructura del proceso a pesar de que le sean suprimidos el 20% de los datos.

En ambos casos, es recomendable para corroborar lo expuesto, experimentar con nuevas aplicaciones.

Referencias Bibliográficas

1. Box, G., Jenkins, G., Reinsel, G. "Time series analysis, forecasting and control" (3ª ed.), Prentice Hall., New Jersey, USA, 1994.
2. Wong, F. "Time series forecasting using backpropagation neural networks". Neurocomputing, Vol. 2, (1991) 147-159.
3. Hill, T., O'Connor, M., Remus, W. "Neural network models for time series forecasts". Management Science, Vol. 42, No. 7 (1996) 1082-1092.
4. Wedding II y Cios. "Time series forecasting by combining RBF network, certainty factors, and the Box-Jenkins model". Neurocomputing, Vol. 10 (1996) 149-168.
5. Faraway, J., Chatfield, C. "Time series forecasting with neural networks: a comparative study using the airline data". Applied Statistics, Vol. 47, No. 2 (1998) 231-250.
6. Bowerman, B., O'Connell, R. "Forecasting and time series: an applied approach" (3ª ed.), Duxbury Press, California, USA, 1993.
7. Gujarati, D. "Econometría" (3ª ed.), McGraw-Hill, Santa Fè de Bogotá, Colombia, 1997.
8. Makridakis, S. y Wheelwright, S. "Manual de técnicas de pronósticos", Limusa Noriega Editores, México, 1994.
9. Colina, E., Rivas, F. "Introducción a la inteligencia artificial", Universidad de Los Andes, Mérida, Venezuela, 1998.
10. Haykin, S. "Neural networks: a comprehensive foundation", Macmillan College Publishing Company, USA, 1994.
11. Hagan, M., Demuth, H., Beale, M. "Neural network design", PWS Publishing Company, USA, 1996.
12. Instituto Nacional de Estadística de España (2001). [On line]. Disponible en: <http://www.ine.es>
13. SPSS, Inc. SPSS 8.0 [Programa de Computación].
14. Martín, Q., Cabero, M., Ardonuy, R. "Paquetes estadísticos SPSS 8.0", Hespérides, España, 1999.
15. Collantes, J.: "Predicción con redes neuronales: comparación con las metodologías de Box y Jenkins", tesis de maestría, Universidad de Los Andes, Mérida, Venezuela, 2001.
16. MathWorks, Inc. MatLab 5.1 [Programa de Computación].
17. SAS Institute, Inc. SAS 3.95 [Programa de Computación].
18. Yamakawa, T. "A Neo Fuzzy Neuron and Its Applications to System Identification and Prediction of Chaotic Behavior". IEEE, Vol. 3 (1994) 383-395.

Recibido el 18 de Junio de 2002

En forma revisada el 13 de Septiembre de 2004