

# Pronóstico de series de tiempo con tendencia y ciclo estacional usando el modelo *airline* y redes neuronales artificiales

Forecasting of time series with trend and seasonal cycle using the  
*airline* model and artificial neural networks

Previsão de séries temporais com tendência e ciclo sazonal, que  
usa o modelo *airline* e redes neurais artificiais

J. D. Velásquez<sup>1</sup>, C. J. Franco<sup>2</sup>

*Recepción: 22-ene-2011/Modificación: 07-may-2012/Aceptación: 10-may-2012*

*Se aceptan comentarios y/o discusiones al artículo*

---

## Resumen

Muchas series de tiempo con tendencia y ciclos estacionales son exitosamente modeladas y pronosticadas usando el modelo *airline* de Box y Jenkins; sin embargo, la presencia de no linealidades en los datos son despreciadas por este modelo. En este artículo, se propone una nueva versión no lineal del modelo *airline*; para esto, se reemplaza la componente lineal de promedios móviles por un perceptrón multicapa. El modelo propuesto es usado para pronosticar dos series de tiempo benchmark; se encontró que el modelo propuesto es capaz de pronosticar las series de tiempo con mayor precisión que otras aproximaciones tradicionales.

**Palabras claves:** predicción, modelos no lineales, SARIMA, perceptrón multicapa.

---

<sup>1</sup> Doctor en Ingeniería-Sistemas Energéticos, jdvelasq@unal.edu.co, Profesor Titular, Universidad Nacional de Colombia.

<sup>2</sup> Doctor en Ingeniería-Recursos Hidráulicos, cjfranco@unal.edu.co, Profesor Asociado, Universidad Nacional de Colombia.

## Abstract

Many time series with trend and seasonal pattern are successfully modeled and forecasted by the *airline* model of Box and Jenkins; however, this model neglects the presence of nonlinearity on data. In this paper, we propose a new nonlinear version of the *airline* model; for this, we replace the moving average linear component by a multilayer perceptron neural network. The proposed model is used for forecasting two benchmark time series; we found that the proposed model is able to forecast the time series with more accuracy than other traditional approaches.

**Key words:** prediction, nonlinear macroeconomics, SARIMA, multilayer perceptrons.

## Resumo

Muitas séries temporais com tendência e sazonalidade são sucesso modelado e previsto pelo modelo *airline* de Box e Jenkins, no entanto, este modelo negligencia a presença de não-linearidade dos dados. Neste trabalho, propomos uma nova versão não-linear do modelo *airline*, por isso, substituir o componente linear das medias moviles por um perceptron multicamadas. O modelo proposto é utilizado para previsão de duas séries temporais de referência; descobrimos que o modelo proposto é capaz de prever a série de tempo com mais precisão que outros métodos tradicionais.

**Palavras chaves:** predição, modelos não-lineares, SARIMA, perceptron multicamadas.

---

## 1 Introducción

Una de las características comunes a muchas series de tiempo es la presencia simultánea de una tendencia de largo plazo y una componente cíclica estacional con un período fijo. La metodología de Box y Jenkins [1] es una de las principales herramientas utilizadas en la práctica para la modelación y el pronóstico de series de tiempo con dichas características. Uno de los ejemplos típicos es la aplicación de la metodología a la serie del total mensual de pasajeros en vuelos internacionales de una aerolínea; para esta serie, en [1] se demuestra que el modelo lineal más adecuado es un  $SARIMA(0, 1, 1) \times (0, 1, 1)_s$ , en el cual, el valor actual de la serie es función de los errores, shocks o perturbaciones pasadas, que se calculan como la diferencia entre el pronóstico y el valor real observado. En la práctica se ha encontrado que dicho modelo,

comúnmente conocido en la literatura como *airline* debido al caso de aplicación, es adecuado para representar la dinámica de muchas series de tiempo que presentan tendencia y ciclo periódico estacional.

No obstante, el modelo  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_s$  es lineal y se sabe que en los datos pueden existir patrones determinísticos que un modelo lineal es incapaz de capturar [2, 3, 4]; de ahí, que técnicas no lineales, tales como las redes neuronales artificiales, hayan ganado popularidad durante las últimas décadas para pronosticar este tipo de series de tiempo [5].

Al analizar la literatura más relevante, y particularmente en los últimos doce años, se encuentra que la mayor parte de los nuevos modelos propuestos pueden ser considerados como una generalización al caso no lineal de los modelos  $\text{SARIMA}(P, 0, 0) \times (0, 0, 0)_s$ . Esto es equivalente a considerar que el modelo tiene únicamente como entradas los últimos  $P$  rezagos de la serie de tiempo, y que no se consideran como entradas los errores o shocks pasados, como si pasa en los modelos SARIMA. Diferentes ideas han sido planteadas en la literatura para obtener nuevos modelos no lineales, basados en redes neuronales artificiales, que usan como información de entrada las observaciones pasadas de la serie de tiempo; por ejemplo, una red neuronal artificial puede ser utilizada para calcular los parámetros de otro modelo más simple: Medeiros y Veiga [6] proponen un nuevo modelo NCSTAR para el pronóstico de series de tiempo, en el cual, las salidas de la red neuronal corresponden a los coeficientes de un modelo lineal; Medeiros y Veiga [7] consideran un modelo lineal con coeficientes cambiantes en el tiempo, los cuales son calculados usando una red neuronal artificial; Inoussa et al. [8] usan un modelo AR dependiente del estado cuyos coeficientes son calculados usando una red neuronal wavelet de pesos funcionales. Igualmente, se han formulado técnicas híbridas que combinan dos o más modelos: Lee et al. [9] combinan una red neuronal de regresión general y una red Fuzzy ART, mientras que Pang et al. [10] combinan un modelo de perturbación lineal y una red neuronal artificial; así mismo, Hassan et al. [11] proponen un modelo híbrido que usa un modelo de Markov, redes neuronales artificiales y algoritmos genéticos para el pronóstico de mercados financieros. Otros autores han desarrollado nuevas arquitecturas o introducido modificaciones en las existentes: Lai y Wong [12] crean una red neuronal estocástica que tiene una arquitectura similar a una red neuronal autorregresiva en que los valores de activación de las neuronas son aleatorios; Chen et al. [13] obtienen un árbol neuronal flexible a partir de

una metodología evolutiva para generar la arquitectura de una red neuronal artificial; Ghiassi y Saidane [14] desarrollan un nuevo tipo de red neuronal con arquitectura dinámica que usa como entradas los valores pasados de la serie de tiempo; Suarez-Farinas et al. [15] obtienen un nuevo modelo basado en una mixtura de modelos lineales estacionarios y no estacionarios que mapean, por tramos, la dinámica no lineal de las series de tiempo; el modelo neuronal adaptativo de Wong et al. [16] usa internamente dos modelos de redes neuronales para realizar el pronóstico.

Sin embargo, el desarrollo de modelos que usan como entradas los errores en periodos pasados para realizar la predicción no ha tenido un desarrollo tan amplio; en la literatura más relevante se han propuesto varios modelos que logran esto mediante la hibridización de los modelos ARIMA y las redes neuronales:

- Zhang [17] propone un modelo híbrido que consta de dos fases: en la primera fase, se desarrolla un modelo ARIMA para capturar la componente lineal de la serie de tiempo; y en la segunda, la componente no lineal es representada mediante un perceptrón multicapa que usa como entradas los residuos del modelo ARIMA. Los parámetros óptimos del modelo ARIMA y de la red neuronal son calculados independientemente. El pronóstico del modelo es obtenido sumando el pronóstico del valor actual de la serie (que es calculado usando el modelo ARIMA) y el pronóstico del error usando la red neuronal artificial.
- Aladag et al. [18] usan la misma aproximación que en [17], pero cambiando el perceptrón multicapa por una red neuronal de Elman.
- Tseng et al. [19] usan un modelo SARIMA para capturar la componente lineal de los datos; luego, aplican una red neuronal artificial tipo perceptrón multicapa que recibe como entradas los pronósticos y los residuos del modelo SARIMA.
- Yu et al. [20] proponen una metodología basada en la combinación de la predicción de un modelo linear autorregresivo generalizado y una red neuronal artificial.
- Zou et al. [21] proponen el uso de la técnica de combinación de pronósticos, la cual es usada para combinar la predicción de una red neuronal

artificial y un modelo ARIMA.

- Khashei y Bijari [22], al igual que en [17], proponen un modelo híbrido estimado en dos etapas. En la primera etapa se estima un modelo ARMA con el fin de obtener la secuencia de residuos o shocks; en la segunda etapa, se usa un perceptrón multicapa que capture la componente no lineal remanente; la red neuronal artificial usa como entradas los valores pasados de la serie de tiempo y los residuos del modelo ARMA.
- Khashei y Bijari [23] usan la misma aproximación de Khashei y Bijari [22], pero, incluyen el pronóstico del modelo ARMA como una entrada adicional a la red neuronal.

Una de las mayores diferencias entre el  $SARIMA(0, 1, 1) \times (0, 1, 1)_s$  y los modelos de redes neuronales artificiales, es que el modelo  $SARIMA(0, 1, 1) \times (0, 1, 1)_s$  utiliza los shocks pasados como entradas, mientras que las redes neuronales artificiales utilizan los valores históricos de la serie de tiempo; de ahí, lo que si se sabe es que el modelo  $SARIMA(0, 1, 1) \times (0, 1, 1)_s$  es adecuado (metodológicamente) para representar la dinámica lineal de una serie de tiempo, pero no se puede generalizar su estructura de forma directa a un modelo no lineal que aproveche las demás componentes determinísticas ignoradas en los datos. Consecuentemente con lo anterior, el objetivo de este artículo es desarrollar una versión no lineal del modelo *airline* basada en modelos de redes neuronales artificiales, la cual difiere notoriamente de las otras propuestas presentadas en la literatura, previamente descritas en los párrafos anteriores.

El resto de este artículo está organizado como sigue: en la Sección 2 se presenta el modelo *airline*; en la Sección 3 se describe la arquitectura de las redes neuronales tipo perceptrón multicapa. Seguidamente, se presenta el modelo propuesto en la Sección 5; posteriormente se presentan dos casos de aplicación. Finalmente, se concluye en la Sección 6.

## 2 Modelo *airline*

El modelo *airline* es descrito por la siguiente expresión matemática:

$$(1 - B^s)(1 - B)y_t = (1 - \theta_1 B)(1 - \theta_{s,1} B^s)e_t \quad (1)$$

donde:

- La secuencia  $y_t$  representa la serie de tiempo.
- La secuencia  $e_t$  son los errores, shocks o perturbaciones del modelo, que se calculan como la diferencia entre el valor real y el valor pronosticado.
- $B$  es el operador de retardo, tal que:  $B^n y_t = y_{t-n}$ .
- $\theta_1, \theta_{s,1}$  son los parámetros del modelo.
- $s$  es el periodo de la componente cíclica.  $s = 12$  para datos mensuales con un periodo cíclico anual.

En la ec. (1), el factor de diferenciación simple  $(1-B)$  remueve la tendencia de largo plazo de la serie, mientras que el factor de diferenciación estacional  $(1-B^s)$  elimina el patrón cíclico estacional repetitivo presente en los datos. La serie resultante,  $(1-B)(1-B^s)y_t$ , es débilmente estacionaria; esto es, su media y su varianza son constantes y no dependen del tiempo.

Los valores óptimos de los parámetros  $\{\theta_1, \theta_{s,1}\}$  son calculados minimizando la sumatoria de  $e_t^2$ . La ec. (1) es equivalente a (con  $s = 12$  para datos mensuales):

$$y_t = e_t - \theta_1 e_{t-1} - \theta_{12,1} e_{t-12} + \theta_1 \theta_{12,1} e_{t-13} + y_{t-1} + y_{t-12} - y_{t-13} \quad (2)$$

### 3 Perceptron Multicapa

Las redes neuronales artificiales, en Inteligencia Artificial, son modelos computacionales que imitan las redes de neuronas del cerebro con el fin de emular el comportamiento inteligente [24]; pero, también se definen como modelos no paramétricos que son equivalentes a modelos no lineales de regresión [25]. Su popularidad en la solución de diferentes problemas prácticos, como la predicción de series de tiempo, es debida a características como su robustez y tolerancia al ruido [26, 27].

La representación pictórica de un perceptrón multicapa es presentada en la Figura 1; los cuadrados representan neuronas de entrada que reciben la entrada y la propagan directamente por la red neuronal; existe una neurona

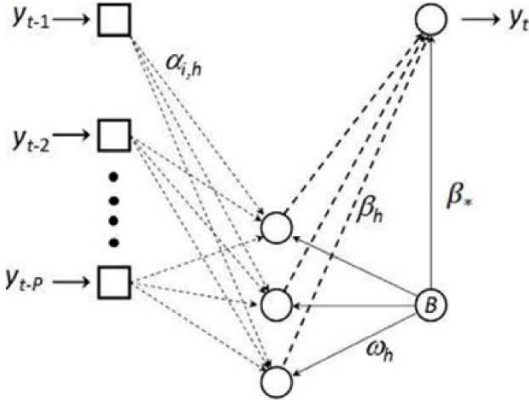
de entrada por cada valor rezagado  $y_{t-i}$ . Los círculos representan neuronas de procesamiento; en ellas, la entrada neta a cada neurona es transformada por la función no lineal  $g()$ ; la neurona de procesamiento marcada con la letra  $B$  es una neurona especial, llamada adaptativa, que no recibe ninguna entrada y siempre produce una salida unitaria positiva. Las flechas representan las conexiones por las que se propagan los datos. Las neuronas están agrupadas por capas; existe una capa de entrada, una o más capas ocultas de procesamiento y una capa de salida con una sola neurona que entrega el valor calculado (pronosticado) para  $y_t$ . La representación matemática de la red neuronal presentada en la Figura 1 es:

$$y_t = \beta_* + \sum_{h=1}^H \beta_h g \left( \omega_h + \sum_{i=1}^p \alpha_{i,h} \times y_{t-i} \right) \quad (3)$$

donde:

- $\beta_*$  es el peso de la conexión de la neurona adaptativa a la neurona de salida.
- $\beta_h$  es el peso asociado a la conexión entre cada neurona  $h$  de la capa oculta y la neurona de salida.
- $\omega_h$  representa el peso de las conexión que van de la neurona adaptativa hasta cada neurona  $h$  de la capa oculta.
- $\alpha_{i,h}$  corresponde a los pesos entre cada entrada  $i$  y la neurona  $h$  en la capa oculta.
- $H$  representa el número de neuronas en la capa oculta.
- $p$  es el número de rezagos considerados.
- $g()$  es la función de activación de las neuronas de procesamiento en la capa oculta.

En la literatura técnica se han propuesto diversas especificaciones para la función  $g()$ ; una de las más comunes es la tangente hiperbólica.



**Figura 1:** Representación pictórica de un perceptrón multicapa para la predicción de series de tiempo.

### 4 Modelo Propuesto

En el caso del modelo *airline*, el factor compuesto  $(1 - B^s)(1 - B)$  permite remover la componente de tendencia de largo plazo y el ciclo periódico estacional de los datos. En [1] se parte de la hipótesis de estacionariedad de los datos; esto es, para una serie débilmente estacionaria, la media y la varianza son independientes del tiempo, hecho que no sucede ante la presencia de tendencia o ciclos estacionales. La tendencia y el ciclo estacional son explicados matemáticamente por la presencia de raíces unitarias simples y estacionales en el modelo que corresponden al coeficiente unitario tácito de  $B$  y  $B^s$  en el factor  $(1 - B^s)(1 - B)$ .

El modelo *airline* utiliza los rezagos (o retardos) 1,  $s$  y  $s + 1$  tanto para los valores pasados de la serie de tiempo  $\{y_t\}$  como para los errores  $\{e_t\}$ , pero sus parámetros  $\{\theta_1, \theta_{s,1}\}$  están asociados a los errores. En el caso del perceptrón multicapa de la ec. (3), se utilizan únicamente los valores pasados de la serie de tiempo y el modelo no incorpora explícitamente los errores pasados; de ahí, que las entradas para ambos modelos son diferentes. Obligando al cumplimiento de ciertas restricciones, Box y Jenkins [1] demostraron que para cada modelo que usa los valores pasados de  $y_t$  existe un modelo equivalente que usa únicamente los valores pasados de  $e_t$ . Así, un modelo que usa como entrada únicamente a  $y_{t-1}$  puede reemplazarse por un modelo que usa como



entradas a la secuencia infinita de errores  $e_{t-1}, e_{t-2}, \dots$  y un modelo que tiene como única entrada a  $e_{t-1}$  puede reemplazarse por un modelo que usa como entradas la secuencia infinita de valores rezagados de la serie de tiempo  $y_{t-1}, y_{t-2}, \dots$ . Bajo esta aseveración, un modelo no lineal debería usar las entradas  $y_{t-1}, y_{t-2}, \dots$  para que sea equivalente, al menos en términos de las variables usadas, al modelo *airline*. Sin embargo, esta alternativa es inadecuada ya que en la práctica debería usarse una secuencia finita pero suficientemente larga de rezagos  $y_{t-i}$  para poder reemplazar cada valor rezagado de  $e_t$ , lo que va en contra de la parsimonia del modelo. Adicionalmente, se aumenta la complejidad en el proceso de estimación de los parámetros debido al incremento en la cantidad de parámetros a ser estimados.

En la aproximación propuesta en este trabajo, se reemplaza la parte derecha de la igualdad en la ec. (1) por el perceptrón multicapa definido en la ec. (3), pero usando como entradas los shocks pasados. Así, el modelo propuesto es:

$$y_t = e_t + y_{t-1} + y_{t-12} - y_{t-13} + \beta_* + \sum_{h=1}^H \beta_h \times g(+\omega_h + \alpha_{1,h}e_{t-1} + \alpha_{2,h}e_{t-12} + \alpha_{3,h}e_{t-13}) \quad (4)$$

Este modelo es notado en este artículo como MA-MLP, debido a que corresponde a una versión no lineal de la componente de promedios móviles (MA, por su sigla en inglés).

Respecto a las otras propuestas presentadas en la literatura más relevante, se debe enfatizar lo siguiente:

- La ec. (1) es similar en estructura a un modelo ARMA sujeto a que algunos coeficientes tienen valor unitario. En [28] se reportó que una red neuronal artificial puede capturar la dinámica de un proceso ARMA y realizar predicciones precisas; según este hallazgo, el modelo presentado en la ec. (4) es un competidor que debe tener, a lo sumo, un desempeño igual al modelo SARIMA.
- Nelson et al. [29] y Zhang y Qi [30] presentan resultados experimentales que favorecen el preprocesamiento de la serie usando operadores de diferenciación simple y estacional [esto es, el factor  $(1 - B)(1 - B^s)$ ],

cuando los datos provienen de un proceso SARIMA. Esta misma recomendación es dada en [26, 27], argumentando que si una serie de tiempo presenta una tendencia y un patrón cíclico estacional, y estos no son removidos, la red neuronal se concentra en aprender dichas componentes, despreciando otros patrones determinísticos en los datos que podrían mejorar la precisión del pronóstico. En el modelo propuesto (ec. (4)), los sumandos  $y_{t+1} + y_{t+12} - y_{t-13}$  son equivalentes a la remoción de las componentes de tendencia y de ciclo estacional por la aplicación del factor  $(1 - B)(1 - B^s)$ .

- En los trabajos de Nelson et al. [29], Zhang y Qi [30] y Masters [26, 27], la red neuronal artificial se construye usando como entradas los valores rezagados de la serie  $w_t = (1 - B)(1 - B^s)y_t$ , mientras que en esta investigación las entradas están conformadas por los shocks pasados. Si efectivamente, la dinámica es representada por un modelo de promedios móviles que usa los shocks pasados como entradas, se requiere la inclusión de muchos rezagos de la serie como entradas para que dicha dinámica sea aproximada por una red neuronal artificial, aumentando el número de parámetros a ser estimados y dificultando el proceso de estimación de los parámetros óptimos. Así, el modelo representado por la ec. (4) debería ser más parsimonioso que un modelo de redes neuronales artificiales que utilice solo los valores pasados de  $y_t$ .
- En los modelos propuestos en [19], [17], [21] y [22] y [23], la estimación de los parámetros es realizada en dos etapas; en la primera etapa, se estiman los parámetros del modelo ARIMA usando técnicas tradicionales; en la segunda etapa, se estiman los parámetros (pesos) de la red neuronal artificial dejando fijos los parámetros del modelo ARIMA. En nuestro trabajo, los parámetros óptimos de la red neuronal artificial  $\{\beta_*, \omega_h, \alpha_{1,h}, \alpha_{2,h}, \alpha_{3,h}; h = 1, \dots, H\}$  son estimados directamente a partir de la serie de datos estudiada, minimizando el error cuadrático medio. Tal como es discutido en la literatura de redes neuronales artificiales, la optimización debe realizarse varias veces usando distintos valores iniciales para los parámetros, y se seleccionan aquellos parámetros que producen el error cuadrático medio más bajo. Como criterio de parada en cada optimización se usa, únicamente, la convergencia a un óptimo local.

## 5 Casos de Aplicación

Para los dos casos de aplicación, se implementaron prototipos del modelo propuesto para cada conjunto de datos y cada configuración considerada. Las implementaciones fueron realizadas en Microsoft Excel, debido, principalmente, a la facilidad para construir el modelo y las funcionalidades del Solver, como lo son: la posibilidad de seleccionar entre dos algoritmos de optimización (GRG con reinicios aleatorios múltiples y Algoritmos Evolutivos) y el manejo de restricciones.

### 5.1 Serie G de Box y Jenkins

La serie G [1] corresponde al número mensual de pasajeros en vuelos aéreos internacionales de una aerolínea entre 1949:1 y 1960:12 (144 observaciones). Dicha serie presenta una tendencia creciente y un comportamiento estacional multiplicativo donde la amplitud del ciclo estacional es directamente proporcional al nivel de la serie [1]. De ahí, que esta serie sea comúnmente utilizada para ejemplificar el uso de los modelos SARIMA y como un benchmark para evaluar el desempeño de modelos no lineales de predicción. Los datos originales de la serie se notan como  $w_t$ .

Usualmente, la serie es transformada usando la función logaritmo natural con el fin de hacer la estacionalidad aditiva. En [1] se indica que el modelo SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> es el más adecuado para representar la dinámica de los datos ( $s = 12$ ) cuando se utiliza el total de las observaciones para el análisis.

Faraway y Chatfield [31] compararon la precisión del pronóstico cuando se usa el modelo *airline* y diferentes perceptrones multicapa que difieren en la cantidad de neuronas en la capa oculta y la combinación de rezagos utilizada como entrada a la red neuronal artificial. Para ello, se usaron los primeros 132 datos para la estimación de los parámetros de los modelos y los 12 datos restantes para evaluar la capacidad predictiva de las metodologías consideradas. Los perceptrones multicapa usados por Faraway y Chatfield [31] son descritos en la ec. (3), con  $g(u) = [1 + \exp(-u)]^{-1}$ . Adicionalmente, las redes neuronales fueron estimadas en [31] para la serie  $z_t = w_t/100$ , y no para  $y_t = \log w_t$ , lo que dificulta el proceso de entrenamiento. Faraway y Chatfield [31]

evalúan la precisión del pronóstico usando la sumatoria del error cuadrático medio (SSE); los valores reportados en el estudio original se reproducen en la Tabla 1; la columna “estimación” corresponde a la SSE para la muestra de estimación de los parámetros de los modelos; “Pronóstico 1P” es la SSE para el pronóstico un mes adelante usando como entradas los datos reales de la serie; “Pronóstico 12P” es la SSE para la muestra de pronóstico en donde se realiza el pronóstico doce meses adelante usando como último dato conocido el último dato de la muestra de entrenamiento; esto es, los shocks pronosticados para los meses anteriores se usan como entradas a la red neuronal para pronosticar el shock actual.

**Tabla 1:** Resumen de los estadísticos de ajuste para la serie G de Box y Jenkins

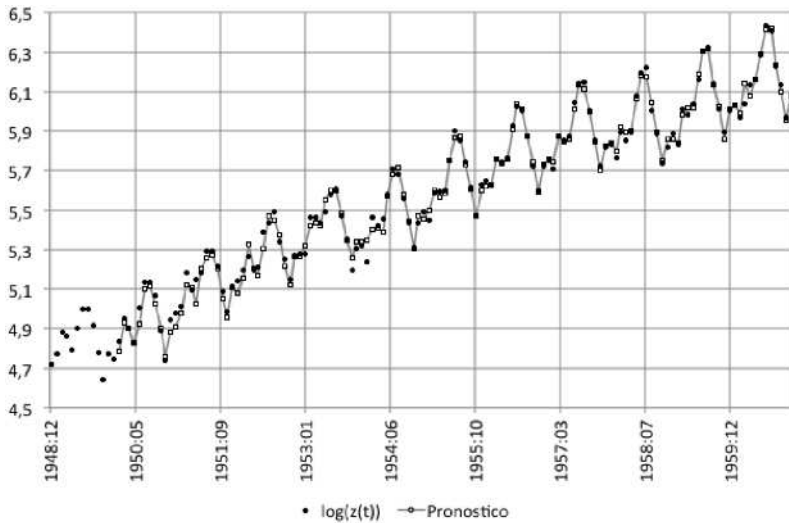
Modelo	Rezagos	H	SSE Estimación 1 P	SSE Pronóstico 1 P	SSE Pronóstico 12 P
SARIMA	1, 12, 13		1,08	0,43	0,39
MLP	1, 2, 3, 4	2	7,74	1,03	58,52
MLP	1-13	2	0,73	0,71	1,08
MLP	1-13	4	* 0,26	1,12	4,12
MLP	1, 12	2	2,30	0,34	0,35
MLP	1, 12	4	2,16	0,44	0,38
MLP	1, 12	10	1,77	0,59	0,51
MLP	1, 2, 12	2	2,17	0,29	0,34
MLP	1, 2, 12	4	1,91	1,03	6,82
MLP	1, 2, 12, 13	2	0,99	0,52	0,37
MLP	1, 2, 12, 13	4	0,81	0,52	0,34
MLP	1, 12, 13	1	1,18	0,50	0,33
MLP	1, 12, 13	2	1,03	0,50	0,33
MLP	1, 12, 13	4	0,84	0,62	0,54
MA-MLP-1	1, 12, 13	1	1,06	0,44	0,48
MA-MLP-2	1, 12, 13	2	0,79	* 0,34	* 0,30
MA-MLP-3	1, 12, 13	3	0,77	0,40	* 0,30

\* Valores mínimos de cada columna. Fuente: [31] y cálculos propios.

Usando los primeros 132 datos para estimar los parámetros del modelo *airline*, se obtiene que  $\theta_1 = 0,3267$  y  $\theta_{12,1} = 0,5777$ . Los parámetros anteriores se obtuvieron al minimizar la suma de  $e_t^2$  para  $t = 14, \dots, 132$ . Es necesario

enfaticar que el modelo *airline* fue estimado para la serie transformada usando la función logaritmo natural (esto es, se modeló la serie  $y_t = \log w_t$ , siendo  $w_t$  la serie original); no obstante, los estadísticos de ajuste presentados en [31] fueron calculados para la serie original y el pronóstico transformados usando  $z_t = w_t/100$ . Así, en la investigación de Faraway y Chatfield [31] se reporta que la sumatoria de errores al cuadrado para el modelo *airline* es de 1,08 para la muestra de calibración, y de 0,43 para la muestra de pronóstico ( $t = 133, \dots, 144$ ). Este cálculo fue verificado en nuestra investigación y se reprodujeron dichos estadísticos de error reportados en [31].

En nuestra investigación, se estimó el modelo propuesto en la ec.(4) con  $H = 1, 2$  y  $3$ . La función de activación utilizada para la capa oculta fue  $g(u) = \tanh u$ . La decisión de usar la función  $g(u) = \tanh u$  está basada únicamente en la facilidad para estimar los parámetros del modelo. Para que los resultados sean comparables al modelo *airline*, la red fue entrenada para la serie  $y_t = \log w_t$ ; no obstante, se usó la misma transformación que en [31] para el cálculo de los errores reportados en la Tabla 1. Se observa que el mejor modelo obtenido en esta investigación usa dos neuronas en la capa oculta (MA-MLP-2), puesto que presenta la SSE más baja para la muestra de predicción; el modelo MA-MLP-3 presenta sobreajuste a los datos de entrenamiento: su ajuste a la muestra de estimación es mejor pero su SSE para la muestra de pronóstico es más alta. El modelo propuesto captura mejor la dinámica de la serie que los modelos SARIMA y MLP de Faraway y Chatfield [31], lo que se evidencia por tener el mejor ajuste a las muestras de estimación y pronóstico. Nótese que el MLP con rezagos 1 – 13 y 2 neuronas en la capa oculta reportado en [31] presenta el menor SSE para la muestra de calibración, pero, esto se debe a que memorizó los datos ya que presenta valores muy altos de la SSE para la muestra de pronóstico. El pronóstico obtenido con el modelo MA-MLP-2 es presentado en la Figura 2.



**Figura 2:** Pronóstico un mes adelante usando el modelo MA-MLP-2. Serie G de Box y Jenkins transformada usando la función logaritmo natural.

## 5.2 Serie Sutter de McCleary y Hay

Esta serie, de 252 observaciones, corresponde a número total de personas empleadas como mano de obra por el condado de Sutter (California) entre 1946:1 y 1966:12. Se caracteriza por la presencia de una tendencia creciente en el largo plazo y un marcado ciclo estacional de periodo anual. Usando la totalidad de los datos, McCleary y Hay [32] identifican el modelo *airline* con parámetros  $\theta_1 = 0,60$  y  $\theta_{12,1} = 0,68$  como el estadísticamente más apropiado para representar la dinámica de los datos.

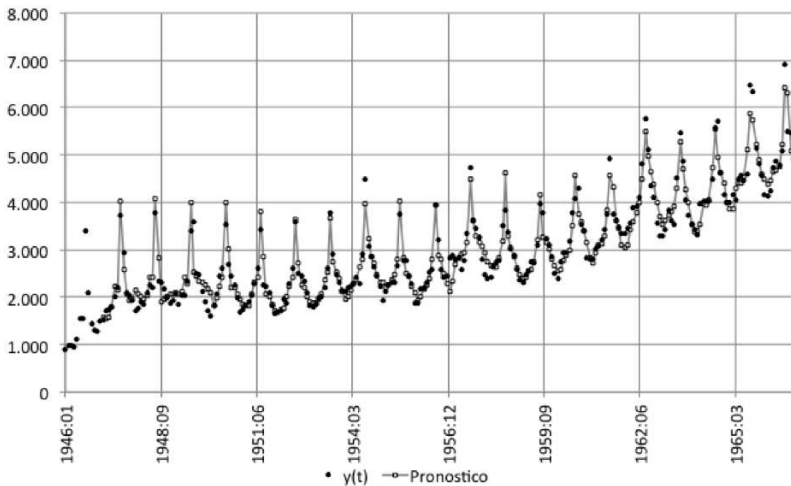
En nuestra investigación, se usaron las primeras 216 observaciones para la estimación de los modelos y las 36 restantes para evaluar la capacidad de pronóstico de las metodologías consideradas. Como estadístico de ajuste se usó el error cuadrático medio (MSE, por su sigla en inglés). El modelo *airline* reestimado con la nueva muestra de calibración tiene como parámetros  $\theta_1 = 0,61$  y  $\theta_{12,1} = 0,68$ , cuyos valores son consistentes con los estimados en [32]. Los estadísticos de ajuste son presentados en la Tabla 2.

**Tabla 2:** Resumen de los estadísticos de ajuste para la serie Sutter de McCleary y Hay.

Modelo	Rezagos	H	SSE Estimación 1 P	SSE Pronóstico 1 P	SSE Pronóstico 12 P
SARIMA	1, 12, 13		59.105	97.513	413.350
MA-MLP-1	1, 12, 13	1	59.015	103.146	171.443
MA-MLP-2	1, 12, 13	2	56.075	* 93.128	* 154.296
MA-MLP-3	1, 12, 13	3	* 55.198	95.228	207.115

\* Valores mínimos de cada columna. Fuente: cálculos propios.

Al igual que el caso anterior, se estimaron modelos con  $H = 1, 2$  y  $3$ . Se encontró que el mejor modelo utiliza dos neuronas en la capa oculta (MA-MLP-2); esta selección está basada en la precisión del modelo para la muestra de pronóstico de 36 observaciones. Adicionalmente, dicho modelo se ajusta de forma más precisa a muestra de entrenamiento que el modelo SARIMA. La predicción un mes adelante es presentada en la Figura 3.



**Figura 3:** Pronóstico un mes adelante usando el modelo MA-MLP-2. Serie Sutter de McCleary y Hay.

## 6 Conclusiones

La presencia de una tendencia de largo plazo y un patrón estacional es una característica común a muchas series de tiempo en economía. En la práctica, se ha encontrado que el modelo *airline* [1] puede capturar adecuadamente la dinámica de un subconjunto de este tipo de series de tiempo y ser efectivo para su pronóstico, siendo así, un modelo catalogado como benchmark en la predicción de series de tiempo. No obstante, dicho modelo es lineal y podrían existir otras componentes ignoradas por el modelo que podrían dar lugar a pronósticos más precisos.

En este artículo, se presenta un nuevo modelo híbrido basado en redes neuronales artificiales que permite capturar la componente no lineal en la serie de tiempo. Dicho modelo es comparado con la aproximación tradicional para dos series de tiempo que son adecuadamente modeladas por el modelo *airline* [1], y se demuestra que el modelo propuesto es capaz de ajustarse mejor y producir pronósticos más precisos que el modelo lineal de Box y Jenkins [1].

Como trabajo futuro se debe investigar sobre el uso de otros tipos de redes neuronales artificiales en vez del perceptrón multicapa utilizado.

## Referencias

- [1] GEP. Box, GM. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day Inc, 1970. Referenciado en 172, 178, 181, 186
- [2] D. van Djck. *Smooth Transition Models: Extensions and Outlier Robust Inference*. PhD thesis, Erasmus University - Rotterdam, 1999. Referenciado en 173
- [3] C. Granger, T. Terasvirta. *Modeling Nonlinear Economic Relationships*. Oxford University Press, 1993. Referenciado en 173
- [4] I. Kaastra, M. Boyd. Designing a neural network for forecasting financial and economic series. *Neurocomputing* **10**: 215–236, 1996. Referenciado en 173
- [5] GP. Zhang, B. Patuwo, M. Hu. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, **14**: 35–62, 1998. Referenciado en 173
- [6] MC. Medeiros, C. Veiga. A hybrid linear-neural model for time series forecasting. *IEEE Transactions on Neural Networks*, **11**(6): 1402–1412, 2000. Referenciado en 173



- 
- [7] MC. Medeiros, A. Veiga. A flexible coefficient smooth transition time series model. *IEEE Transactions on Neural Networks*, **16**(1): 97–113, 2005. Referenciado en 173
- [8] G. Inoussa, H. Peng, J. Wu. Nonlinear time series modeling and prediction using functional weights wavelet neural network-based state-dependent AR model. *Neurocomputing*, **86**(1): 59–74, 2012. Referenciado en 173
- [9] EWM Lee, RKK. Yuen, SM. Lo, KC. Lam, GH. Yeoh. A novel artificial neural network fire model for prediction of thermal interface location in single compartment fire. *Fire Safety Journal*, **39**(1): 67–87, 2004. Referenciado en 173
- [10] B. Pang, S. Guo, L. Xiong, L. Chaoqun. A nonlinear perturbation model based on artificial neural network. *Journal of Hydrology*, **222**(2-4): 504–516, 2007. Referenciado en 173
- [11] R. Hassan, B. Nath, M. Kirley. A fusion model of HMM, ANN and GA for stock market forecasting. *Expert Systems with Applications*, **33**(1): 171–180, 2007. Referenciado en 173
- [12] TL. Lai, SPS. Wong. Stochastic neural networks with applications to nonlinear time series. *Journal of the American Statistical Association*, **96**(455): 968–981, 2001. Referenciado en 173
- [13] Y. Chen, B. Yang, J. Dong, A. Abraham. Time-series forecasting using flexible neural tree model. *Information Sciences*, **174**(3-4): 219–235, 2005. Referenciado en 173
- [14] M. Ghiassi, H. Saidane. A dynamic architecture for artificial neural networks. *Neurocomputing*, **63**: 397–413, 2005. Referenciado en 174
- [15] M. Suarez-Farinas, CE. Pedreira, MC. Medeiros. Local global neural networks: a new approach for nonlinear time series modeling. *Journal of the American Statistical Association*, **99**(468): 1092–1107, 2004. Referenciado en 174
- [16] WK. Wong, M. Xia, WC. Chu. Adaptive neural network model for time-series forecasting. *European Journal of Operational Research*, **207**(2): 807–816, 2010. Referenciado en 174
- [17] GP. Zhang. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, **50**: 159–175, 2003. Referenciado en 174, 175, 180
- [18] CH. Aladag, E. Egrioglu, C. Kadilar. Forecasting nonlinear time series with a hybrid methodology. *Applied Mathematics Letters*, **22**(9): 1467–1470, 2009. Referenciado en 174

- [19] F-M. Tseng, H-C. Yu, GH. Tzeng. Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting and Social Change*, **69**(1): 71–87, 2002. Referenciado en 174, 180
- [20] L. Yu, S. Wang, KK. Lai. A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Computers & Operations Research*, **32**(10): 2523–2541, 2005. Referenciado en 174
- [21] HF. Zou, GP. Xia, FT. Yang, HY. Wang. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing*, **70**(16-18): 2913–2923, 2007. Referenciado en 174, 180
- [22] M. Khashei, M. Bijari. An artificial neural network (p, d, q) model for time series forecasting. *Expert Systems with Applications*, **37**(1): 479–489, 2010. Referenciado en 175, 180
- [23] M. Khashei, M. Bijari. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, **11**(2): 2664–2675, 2011. Referenciado en 175, 180
- [24] N. Kasabov. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. 2nd edition. Massachusetts Institute of Technology, 1998. Referenciado en 176
- [25] W. Sarle. Neural networks and statistical models. *Proceedings of the 19th Annual SAS Users Group Int. Conference*, 1538–1550, 1994. Referenciado en 176
- [26] T. Masters. *Practical Neural Network Recipes in C++*. First edn, Academic Press, Inc, 1993. Referenciado en 176, 180
- [27] T. Masters. *Neural, Novel and Hybrid Algorithms for Time Series Prediction*. First edn, John Wiley and Sons, Inc., 1995. Referenciado en 176, 180
- [28] GP. Zhang. An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research*, **28**(12): 1183–1202, 2001. Referenciado en 179
- [29] M. Nelson, T. Hill, W. Remus, and M. O'Connor. Time series forecasting using neural networks: should the data be deseasonalized first? *Journal of Forecasting*, **18**: 359–367, 1999. Referenciado en 179, 180
- [30] GP. Zhang, M. Qi. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, **160**(2): 501–514, 2005. Referenciado en 179, 180

- [31] J. Faraway, C. Chatfield. Time series forecasting with neural networks: A comparative study using the *airline* data. *Applied Statistics*, **47**(2): 231–250, 1998. Referenciado en 181, 182, 183
- [32] R. McCleary, RA. Hay, Jr. *Applied time series analysis for the social sciences*. Sage Publications, London, 1980. Referenciado en 184