

# Una comparación entre los modelos de componentes no observables y las redes neuronales artificiales

## A comparison between unobserved components models and artificial neural networks

Natalia Castaño, Ing (c); Olga L. Montoya, Ing (c); Juan D. Velásquez. Ph.D.  
Grupo de Computación Aplicada. Universidad Nacional de Colombia Sede Medellín, Colombia  
{ncastanj, olmontoy, jdvelasq}@unal.edu.co

Recibido para revisión 30 de noviembre de 2008, aceptado 2 de abril de 2009, versión final 28 de mayo de 2009

**Resumen-** Las redes neuronales artificiales han sido comúnmente usadas para pronosticar series de tiempo no lineales debido a sus ventajas; sin embargo, el procedimiento de pronóstico depende de heurísticas y juicio experto. La metodología de modelado de componentes no observables es una técnica de pronóstico basada en el filtro de Kalman y la representación de las características visuales de la serie analizada. En este artículo, se compara la habilidad de pronóstico de la aproximación ARIMA, las redes neuronales artificiales y los modelos de componentes no observables utilizando dos series benchmark. Se encontró que el modelo de componentes no observables es muy bueno para una de estas series, mientras que es muy malo para las otras. Sin embargo, es posible concluir que la aproximación analizada es competitiva con las metodologías propias de la inteligencia artificial.

**Palabras Clave-** Sistemas Inteligentes, Agentes Inteligentes y Sistemas Neuro-Difusos, Modelos de Componentes No Observables, Series de Tiempo, Predicción.

**Abstract-** Artificial neural networks have been commonly used to forecast nonlinear time series due to its advantages; however, the forecast procedure depends on heuristics and expert judgment. Unobserved components modeling methodology is a forecast technique based on the Kalman filter and the representation of visual characteristics of the analyzed time series. In this paper, we compare the forecast ability of the ARIMA approach, artificial neural networks and the unobserved components model when two benchmark time series are used. We found that accuracy of the unobserved components model is very good for one of these time series, while is very bad for the other. However, is possible to conclude that the analyzed approach is a competitive with artificial intelligence methodologies.

**Keywords-** Intelligent Systems, Intelligent Agents and Neurofuzzy Systems, Unobserved Components Models, Time Series, Forecasting.

### I. INTRODUCCIÓN

La predicción de series de tiempo es un importante problema de investigación debido a sus dificultades teóricas y sus implicaciones prácticas [1]. En el problema de predicción busca identificar el modelo que permita pronosticar con mayor precisión la evolución futura de la variable estudiada durante uno o más periodos hacia adelante basándose en sus datos rezagados. Su aplicación abarca áreas como la ingeniería, la economía y las ciencias sociales [2]. Tal como es discutido en [2], el problema de la predicción de series temporales ha sido abordado con diferentes metodologías que se agrupan en:

- Métodos estadísticos y econométricos
- Sistemas basados en reglas
- Redes neuronales artificiales y sistemas neurodifusos.
- Modelos híbridos.

Dentro de estas técnicas, las redes neuronales artificiales han gozado de bastante popularidad cuando la serie analizada presenta características no lineales. Esto se debe a que esta clase de modelos pueden aproximar cualquier función continua definida en un dominio compacto [3] [4] [5], y son tolerantes ante el ruido, información incompleta, e inclusive datos contradictorios [6] [7]. Sin embargo, su especificación se basa fundamentalmente en criterios heurísticos y juicio experto del modelador [6] [7], de tal manera, que dicho proceso se basa en un conjunto de pasos críticos que afectan el desempeño final del modelo [8], en términos de su ajuste a los datos históricos y su capacidad de generalización con el fin de realizar pronósticos.

Particularmente, muchos de estos criterios heurísticos están

relacionados con la forma de preparar los datos para realizar el aprendizaje por parte de la red neuronal utilizada. Por ejemplo, en [6] [7] se enfatiza la transformación de la serie utilizando operadores de diferenciación simple y estacional que son tradicionalmente utilizados cuando se usan modelos no lineales [9]. El uso de estos operadores tiene como fin remover características explícitas de la serie, tal como la tendencia o la presencia de patrones cíclicos repetitivos, de tal forma, que la red se concentre en capturar patrones más sutiles que están ocultos en los datos. Sin embargo, este punto no está resuelto satisfactoriamente en la literatura.

Los problemas anteriores motivan la comparación de las redes neuronales con otros modelos alternativos en términos de su facilidad de uso y calidad de sus resultados. Particularmente, dentro de la estadística y la econometría, los modelos de componentes no observables (UCM, por su sigla en inglés) han ganado popularidad debido a su simpleza, a su capacidad de adaptación, y a su elegancia conceptual. Su desarrollo conceptual está bien elaborado, y existen abundantes aplicaciones al pronóstico de series de tiempo [10] [11]. Sin embargo, no existen comparaciones entre la calidad de los pronósticos obtenidos usando UCM y redes neuronales artificiales.

Consecuentemente, el objetivo de este artículo es presentar una comparación entre la habilidad de predicción de los modelos de redes neuronales y los UCM al pronosticar dos series benchmark que ya han sido ampliamente analizadas en la literatura.

En resto de este artículo está organizado como sigue: en la Sección II, se discuten los modelos comparados; seguidamente, en la Sección III se presentan los resultados obtenidos y se realiza su discusión. Finalmente, en la Sección IV se concluye.

## II. MODELOS COMPARADOS

### A. ARIMA

La metodología ARIMA fue introducida originalmente por Box y Jenkins [9] como un medio sistemático y ordenado para modelar series de tiempo. Un modelo ARIMA (p,d,q) es descrito como:

$$(1-B^d) \varphi(B) y_t = \mu + \theta(B) e_t$$

Donde:  $B$  es el operador de rezago, tal que,  $B y_t = y_{t-1}$ .  $\varphi(B)$  es el polinomio  $1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$ .  $\theta(B)$  es el polinomio  $1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ .  $\mu$  es una constante. Este modelo postula que el valor actual de la serie  $y_t$  puede ser obtenido como una combinación lineal de sus valores y errores pasados.

### B. Redes neuronales artificiales

Un perceptrón multicapa (MLP) es un estimador de regresión semi-paramétrico [12], el cual puede aproximar cualquier función continua definida en un dominio compacto, con un nivel de

precisión arbitrario previamente definido [4] [3] [5]. En un perceptrón multicapa, la variable dependiente  $y_t$  se obtiene como:

$$y_t = \beta_* + \sum_{h=1}^H \beta_h G \left( \alpha_{*,h} + \sum_{p=1}^P \alpha_{p,h} y_{t-p} \right) + e_t$$

Con  $e_t = \sigma \epsilon_t$ .  $G()$  es la función de activación de las neuronas de la capa oculta.  $\sigma$  es la desviación estándar de los errores.  $\epsilon_t$  es una variable aleatoria que sigue una distribución normal estándar.  $H$  es el número de neuronas en la capa oculta;  $P$  es el número de rezagos considerados para  $y_t$ . La representación gráfica del modelo anterior se presenta en la Fig. 1.

### C. Modelos de componentes no observables

Los modelos estructurales o de componentes no observables [10] se basan en capturar las características explícitas más sobresalientes de la serie temporal estudiada, que corresponden a sus movimientos de largo plazo y sus patrones cíclicos y repetitivos, los cuales son representados respectivamente por la tendencia y el patrón estacional. En estos modelos, dichas componentes pueden variar estocásticamente en el tiempo, de tal forma que pueden adaptarse a los cambios que sufre la serie. Dichas componentes son:

- Tendencia.
- Ciclos de largo plazo.
- Componente estacional: corresponde a un patrón repetitivo, usualmente de periodo anual.
- Componente irregular: corresponde a una variable aleatoria que representa todas aquellas variables que afectan la serie de tiempo para las cuales no hay datos.

Al combinar estas componentes se obtienen diferentes modelos, los cuales son descritos a continuación.

En el modelo de tendencia local lineal, se asume que el comportamiento de la serie,  $y_t$ , está determinado por una tendencia estocástica,  $\mu_t$ , y una componente irregular,  $v_t$ , tal que su comportamiento puede ser descrito como:

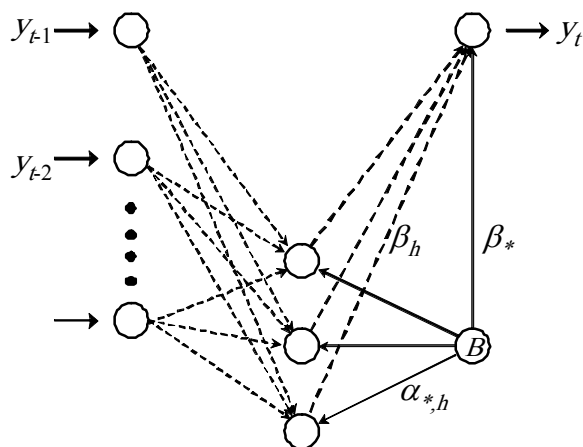


Figura 1. Arquitectura de una red neuronal tipo perceptrón multicapa

$$y_t = \mu_t + v_t, \quad v_t \sim N[0, \sigma_v^2]$$

La serie es perturbada en su nivel y en su pendiente,  $\beta_t$ , de forma aleatoria, tal que:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim N[0, \sigma_\eta^2] \\ \beta_t &= \beta_{t-1} + \chi_t, \quad \chi_t \sim N[0, \sigma_\chi^2] \end{aligned}$$

donde la notación  $N[0, \sigma^2]$  significa que las perturbaciones son independientes e idénticamente distribuidas siguiendo una normal con media cero y varianza  $\sigma^2$ . Cuando las varianzas  $\sigma_\eta^2$  y  $\sigma_\chi^2$  son cero, la tendencia es determinística, y las observaciones oscilan alrededor de ella siguiendo una distribución normal con media cero y varianza  $\sigma_v^2$ . Cuando  $\sigma_\chi^2 = 0$  y  $\sigma_\eta^2 > 0$ , la pendiente se hace constante y el modelo se reduce a un paseo aleatorio. En el caso contrario, cuando  $\sigma_\chi^2 > 0$  y  $\sigma_\eta^2 = 0$  el modelo equivale a un paseo aleatorio integrado.

Cuando la serie estudiada presenta una componente periódica estacional, este patrón de comportamiento se incorpora dentro del modelo a través de un conjunto de variables ficticias  $\gamma_p$  cuya suma sobre un periodo del patrón es cero; el patrón estacional se puede hacer estocástico considerando que la componente actual  $\gamma_t$  es perturbada por un ruido aleatorio  $\omega_t$ , tal que su evolución puede ser descrita como:

$$\gamma_t = \omega_t + \sum_{i=1}^{p-1} \gamma_{t-i}, \quad \omega_t \sim N[0, \sigma_\omega^2]$$

así, el valor actual de la serie se obtiene como:

$$y_t = \mu_t + \gamma_t + v_t, \quad v_t \sim N[0, \sigma_v^2]$$

El proceso de extracción de las componentes no observables  $\mu_t$ ,  $\beta_t$  y  $\gamma_t$  se realiza a partir de la aplicación del filtro de Kalman basada en la información disponible en el momento  $t$ . La estimación de los hiperparámetros del modelo ( $\sigma_v^2$ ,  $\sigma_\eta^2$ ,  $\sigma_\omega^2$  y  $\sigma_\chi^2$ ) se realiza maximizando la función de verosimilitud de los residuales de la predicción un periodo adelante. Los detalles son presentados en [10].

Cuando existe una componente periódica estacional, este patrón de comportamiento se incorpora dentro del modelo ya descrito agregando la componente:

$$\omega_t = \sum_{i=1}^{s/2} \omega_{j,t} \quad \omega_t \sim N[0, \sigma_\omega^2]$$

Donde  $s$  es un ciclo con frecuencia  $x_j = 2\pi j/2$ . Así, el valor actual de la serie se obtiene como:

$$y_t = \mu_t + \omega_t + u_t + o_t, \quad o_t \sim N[0, \sigma_o^2] \quad (1)$$

En el modelo tradicional autorregresivo no se limita a ser un modelo estacionario, se compone de una componente irregular,  $v_t$  [13]

$$u_t = \rho \gamma_{t-1} + v_t \quad v_t \sim N[0, \sigma_v^2]$$

y rezagos específicos de la variable dependiente incluida como predictora [13]:

$$\beta_t = \sum \phi_i^* y_{t-i}$$

tal que su comportamiento puede ser descrito como:

$$y_t = u_t + \beta_t \quad (2)$$

### III. RESULTADOS OBTENIDOS Y DISCUSIÓN

#### A. *Serie PAPER*

La serie a continuación estudia las ventas de la industria de papel para impresión y para escritura entre enero de 1963 y diciembre de 1972 (120 observaciones). Para esta serie se utilizaron los primeros 100 datos para entrenamiento y los 20 restantes para pronóstico [14]. Makridakis et al utilizó el modelo ARIMA (0,1,1)(1,1,1)<sub>12</sub> como el mejor de entre 14 modelos que tomaron como las mejores opciones posibles. El resultado para dicho modelo es reportado en la tabla I.

En [15], esta misma serie es pronosticada usando un perceptrón multicapa. Los investigadores usaron el sistema Clementine del software SPSS. En [15], la mejor red neuronal utiliza los rezagos 1, 7 y 12; los estadísticos de ajuste son reportados, igualmente, en la tabla I.

Para esta investigación uso el procedimiento UCM implementado en el software SAS. Se usaron los mismos conjuntos de valores para el entrenamiento (100 observaciones) y para el pronóstico (20 observaciones), utilizados en otras investigaciones en que se usaron los modelos ARIMA y ANN. Nuestros resultados nos han brindado un modelo con mejores valores para MSE y MAD tanto en entrenamiento como en pronóstico.

El modelo utilizado para pronosticar esta serie es descrito en la ecuación (1), y contiene: una componente de tendencia, una componente estacional y una componente irregular.

Los resultados obtenidos muestran que el UCM es mejor y más apropiado para pronosticar esta serie. Los estadísticos de ajuste son reportados en la Tabla I. Las mejores sobre el modelo ARIMA fueron del 0.9% en entrenamiento y del 20.5% en pronóstico. La predicción obtenida con el modelo UCM es presentada en la Fig. 2.

#### B. *Serie INTERNET*

Esta serie corresponde al número de usuarios registrados en un servidor de Internet durante 100 segundos. Las primeras

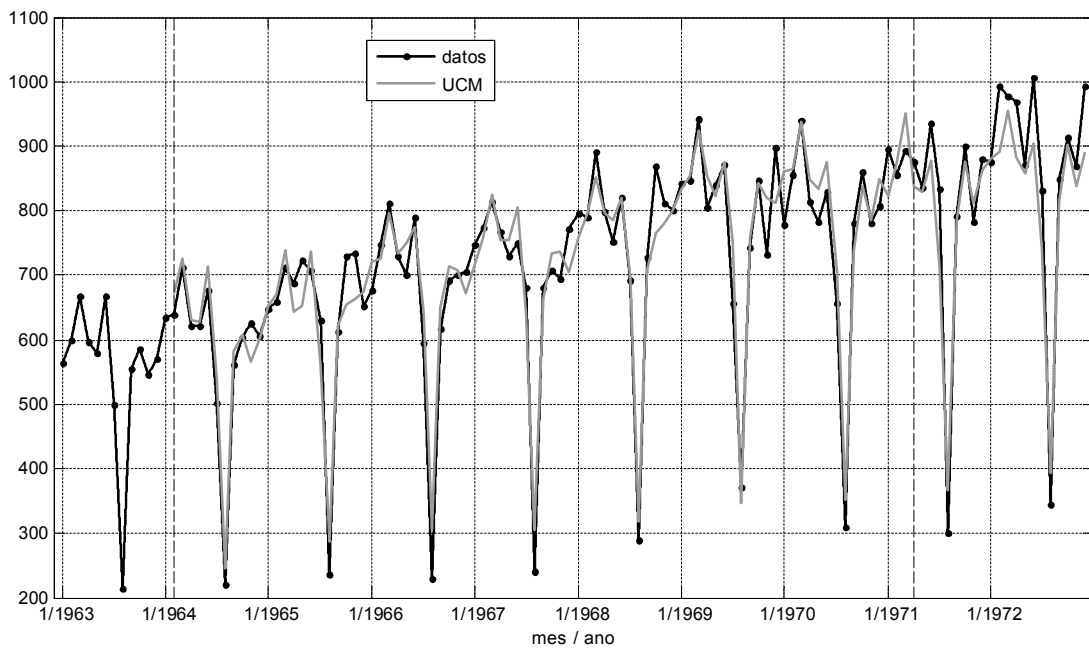


Figura 2. Pronósticos obtenidos para la serie PAPER.

TABLA I. Estadísticos de ajuste para la serie PAPER

Modelo	Entrenamiento MSE (MAD)	Pronóstico MSE (MAD)
ARIMA (1,12,13)	1715.5 (34.04)	4791.7 (54.72)
ANN (1,7,12)	1951.5 (36.79)	5874.8 (62.62)
UCM	1700.8 (2.92)	3809.1 (48.81)

80 observaciones fueron utilizadas para estimar el modelo y las 20 restantes fueron empleados para predecirla. Los datos utilizados en el análisis de la serie no fueron transformados. Esta serie fue modelada originalmente en [14], pero ha sido utilizada en otras investigaciones como un benchmark para la comparación de métodos de predicción. En [14], se consideró que la serie podía ser pronosticada usando un modelo ARIMA; el mejor modelo de predicción es un ARIMA(3,1,0), el cual fue seleccionado entre varias configuraciones alternativas. En [15], se reportan los resultados obtenidos al pronosticar esta serie con una red neuronal tipo perceptrón multicapa que fue obtenida usando el sistema Clementine del software SPSS. Los estadísticos de ajuste a las muestras de entrenamiento y calibración para ambos modelos son reportados en la Tabla II.

Al igual que los estudios anteriores de esta serie, se utilizan los mismos conjuntos de datos para la estimación de los parámetros del modelo y para la predicción de la serie. Se utiliza como criterio de evaluación el error cuadrático medio (MSE) y la desviación absoluta media (MAD). Los resultados presentados en la tabla II muestran que el modelo UCM es mejor que el ARIMA (3,1,0); las mejoras sobre este modelo son del 20.7% en entrenamiento. No obstante, esta aproximación (UCM) presenta un error de entrenamiento peor que el reportado

para el modelo de redes neuronales. Igualmente, el error de pronóstico para el UCM es muy superior a los reportados para los modelos ARIMA y de redes neuronales artificiales. La predicción obtenida usando UCM es presentada en la Fig. 3.

Este modelo es descrito por la ecuación (2).

#### IV. CONCLUSIONES

TABLA II. ESTADÍSTICOS DE AJUSTE PARA LA SERIE INTERNET

Modelo	Entrenamiento MSE (MAD)	Pronóstico MSE (MAD)
ARIMA (3,1,0)	9.76 (2.42)	8.11 (2.23)
ANN	7.00 (2.10)	9.25 (2.25)
UCM (1,2,3,4)	7.72 (1.92)	32.34 (4.88)

En este artículo se comparó el desempeño de la técnica de componentes no observables con los modelos ARIMA y de redes neuronales artificiales al pronosticar dos series benchmark. Se encontró que, en algunos casos, los UCM pueden presentar mejores resultados que las otras dos técnicas consideradas. Sin embargo, las evidencias encontradas no permiten concluir que la aproximación analizada sea consistentemente mejor o peor que las otras técnicas.

Como trabajo futuro, se requiere incrementar el número de series benchmark consideradas para esclarecer la pregunta de investigación objeto de este artículo.

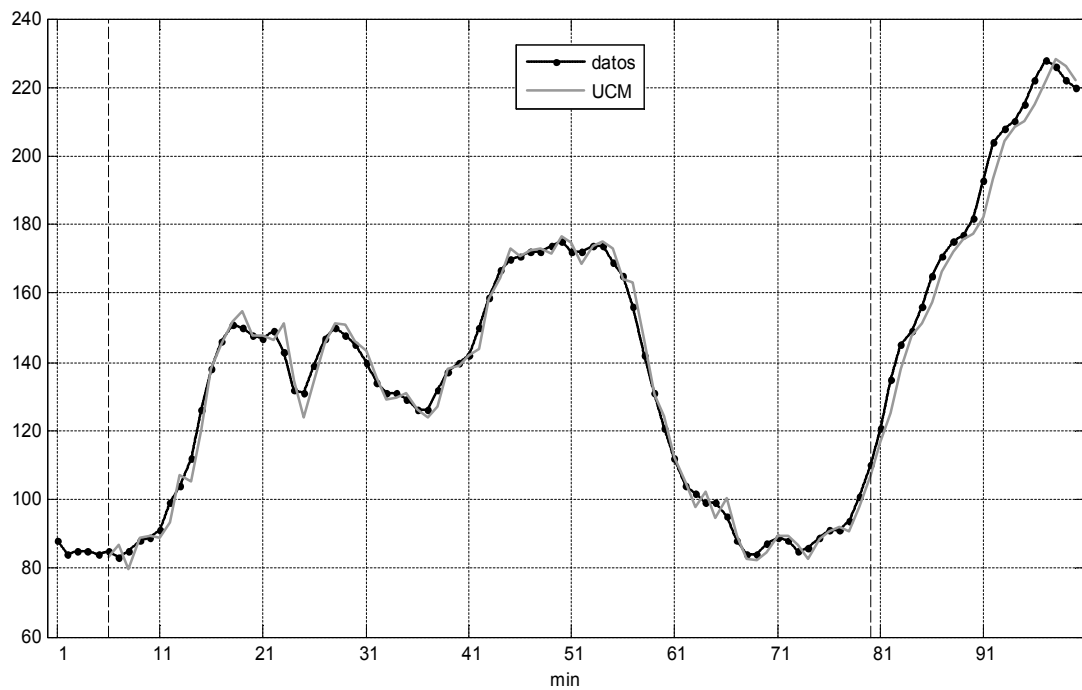
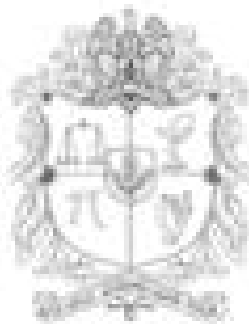
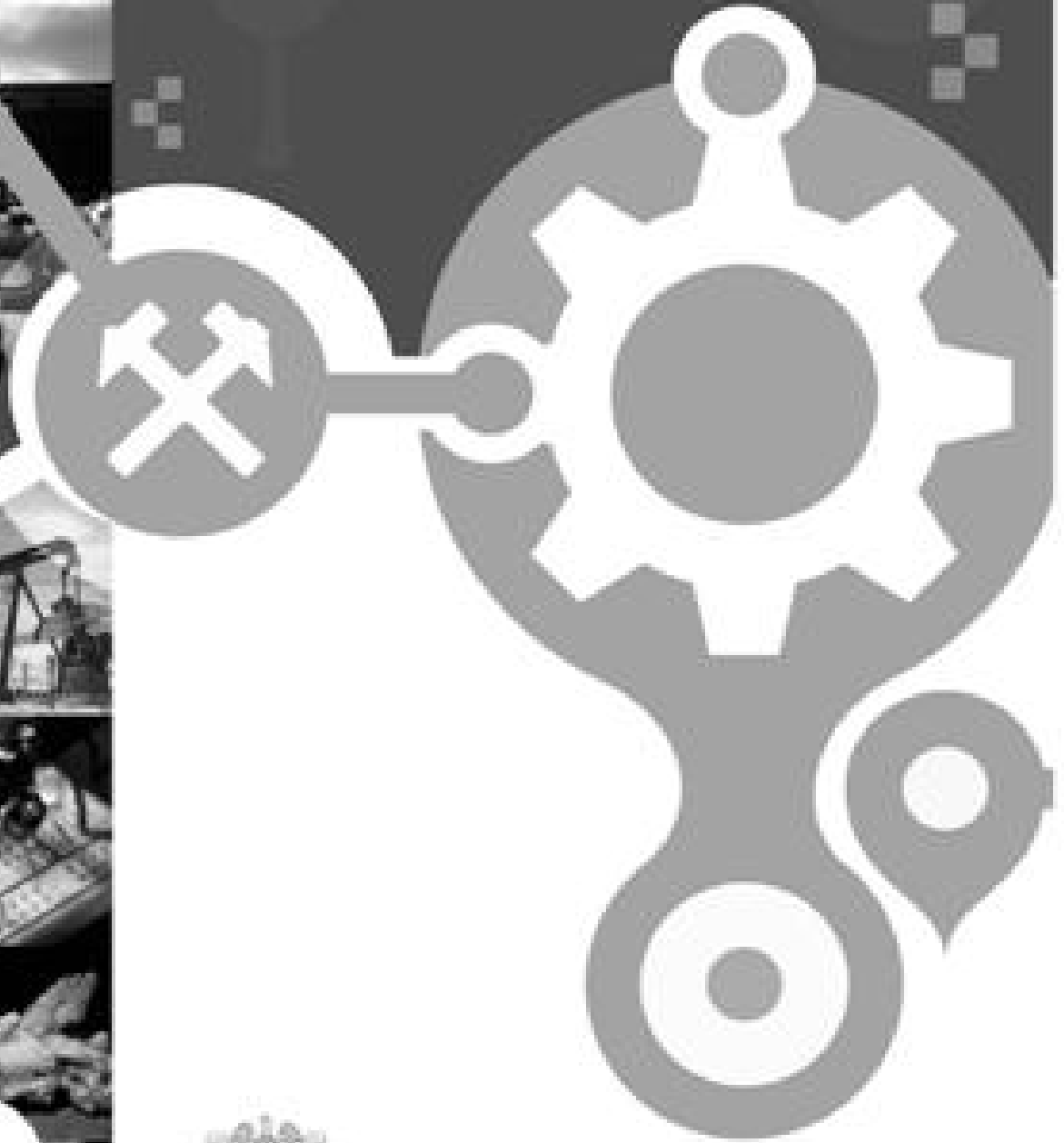


Figura 3. Pronóstico obtenido para la serie INTERNET

#### REFERENCIAS

- [1] Weigend A, & Gersfeld, N. 1993. Time-Series Prediction: Forecasting the future and understanding the past. Reading, Mass, Addison-Wesley.
- [2] Kasabov, N. (1998), Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering, second edn, Massachusetts Institute of Technology.
- [3] Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function, Mathematics of Control: Signals and Systems. 2, (1989), 202-314.
- [4] Hornik, K., Stinchcombe, M. y White, H. (1989), 'Multilayer feedforward networks are universal approximators'
- [5] Funahashi, K. 1989. On the approximate realization of continuous mappings by neural networks, Neural Networks. 2, (1989), 183-192.
- [6] Masters, T. Practical neural network recipes in C++. New York: Academic Press. 1993.
- [7] Masters, T. Neural, Novel and Hybrid Algorithms for Time Series Prediction, first edn, John Wiley and Sons, Inc. 1995
- [8] I. Kaastra, and M. Boyd. "Designing a neural network for forecasting financial and economic series", Neurocomputing , 1996, 10, pp. 215-236.
- [9] Box, G. E. P. y Jenkins, G. M. (1970), Time Series Analysis: Forecasting and Control, Holden-Day Inc.
- [10] Harvey, A. (1989), Forecasting, structural time series models and the Kalman filter, Cambridge University Press
- [11] Durbin, J. and Koopman, S.J. (2001), Time Series Analysis by State Space Methods, Oxford: Oxford University Press.
- [12] Sarle, W. 1994. Neural networks and statistical models. The 19th Annual SAS Users Group Int. Conference. Cary, NC: SAS Institute. 1994, 1538-1550.
- [13] SAS Institute Inc. 2004. SAS/ETS\_ 9.1 User's Guide. Cary, NC: SAS Institute Inc.
- [14] S.G. Makridakis, S.C. Wheelwright y R.J. Hyndman. "Forecasting: Methods and applications". 3rd edition. New York. John Wiley & Sons. 1998.
- [15] Ghiassi, M., H. Saidane, and D. K. Zimbra (2005), "A dynamic artificial

neural network model for forecasting time series events", International Journal of Forecasting 21: 341- 362.



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA  

---

SEDE MEDELLÍN  
FACULTAD DE MINAS