



Comparación de pronósticos de caudales máximos con modelos ARIMA y redes neuronales artificiales

Comparison of maximum flow forecasts with ARIMA models and artificial neural networks

¹ Noelia P. Rodríguez Paredes

ORCID: 0000-0002-3640-2779

² Elmis J. García Zare

ORCID: 0000-0003-4863-7230

³ Luis A. López Puycan

ORCID: 0000-0003-4454-6737

RESUMEN

El presente estudio tuvo como principal objetivo comparar dos técnicas estadísticas de pronóstico en el modelamiento del caudal máximo del río Huaura. La investigación fue de tipo aplicada y longitudinal porque se estudió el comportamiento de los caudales máximos del río Huaura a través del tiempo. Para cumplir con el objetivo se modeló la serie hidrológica mediante modelos ARIMA, donde 611 meses (90 %) de los datos se usó para ajuste del modelo y 68 meses (10 %), para el pronóstico; y mediante redes neuronales artificiales autorregresivas (ARNN), donde 407 meses (60 %) de los datos se usó para entrenamiento, 204 meses (30 %), para prueba y 68 meses (10 %), para reserva del modelo. Finalmente se encontró que el modelo ARNN es más eficiente que los modelos ARIMA para realizar pronósticos.

Palabras claves: Caudal máximo, modelos ARIMA, redes neuronales.

ABSTRACT

The aim of this research was to compare two statistical forecasting techniques in the modeling of the maximum flow of the Huaura River. The study is of an applied and longitudinal type because the behavior of the maximum flows of the Huaura River over time was studied. To achieve the goal, the hydrological series was modeled using ARIMA models, where 611 months (90 %) of the data was used for model adjustment and 68 months (10 %), for the forecast; and through autoregressive artificial neural networks (ARNN) where 407 months (60 %) of the data was used for training, 204 months (30 %), for testing and 68 months (10 %), for model reserve. Finally, it was found that the RNN model is more efficient than the ARIMA models for forecasting.

Keywords: Maximum flow, ARIMA models, artificial neural networks.

¹ Universidad Nacional de Trujillo. Trujillo, Perú. E-mail: nrodriguezp@unitru.edu.pe

² Universidad Nacional de Trujillo. Trujillo, Perú. E-mail: egarciaz@unitru.edu.pe

³ Universidad Nacional Jorge Basadre Grohmann. Tacna, Perú. E-mail: llopezp@unjbg.edu.pe

INTRODUCCIÓN

Las inundaciones suponen un riesgo para las personas y son producidas por ríos, estuarios, la acción del mar o fuertes precipitaciones, causando significativos costes económicos. El último fenómeno del niño trajo como consecuencia el desborde del río Huaura, causando numerosos destrozos, desastres de infraestructura pública y privada, colapso de vías principales de transporte, inundación de viviendas, afectaciones al sistema de alcantarillado y limitando el acceso al agua potable con el consecuente riesgo en la salud de los pobladores. Aunque, los municipios provinciales y distritales tienen establecidos sistemas de prevención contra desastres naturales y planes de acción durante estos eventos, estos procedimientos deben ser complementados con estudios científicos, basados en datos de mediciones que permitan predecir o generar sistemas de aviso ante estos eventos, de esta forma, se puede planificar o tomar medidas no estructurales que pueden ser de gran importancia para la reducción de riesgos latentes. Es común, encontrar investigaciones con propósitos de predecir el comportamiento de ciertas variables, por ejemplo, el aumento de los caudales al paso del tiempo.

Los métodos estadísticos clásicos como los modelos ARIMA son utilizados para tales fines, sin embargo, metodologías más actuales como las redes neuronales artificiales autorregresivas (ARNN) pueden ser una alternativa ante las dificultades que puede significar la gran cantidad de datos a modelar. Esta investigación está basada en utilizar dos técnicas de pronóstico: series de tiempo mediante los modelos ARIMA y ARNN, para estudiar el comportamiento hidrológico y estimar modelos de los caudales máximos del río Huaura; con el fin de poder comparar y establecer qué modelo es el más eficiente al momento de realizar pronósticos de sus valores.

Las redes neuronales autorregresivas (ARNN) se pueden obtener combinando un modelo lineal autorregresivo con la arquitectura de un perceptrón multicapa (MLP). En un modelo de redes neuronales autorregresivas, la variable respuesta y_t es calculada como una función no lineal de sus P valores rezagados y_{t-p} , para $p = 1, \dots, P$:

$$y_t^* = \eta + \sum_{p=1}^P \varphi_p y_{t-p} + \sum_{h=1}^H \beta_h G(\omega_h + \sum_{p=1}^P \alpha_{p,h} y_{t-p})$$

Donde: $G(\cdot)$ es la función tangente hiperbólica definida como:

$$G(u) = \frac{2}{1 + e^{-2u}} - 1$$

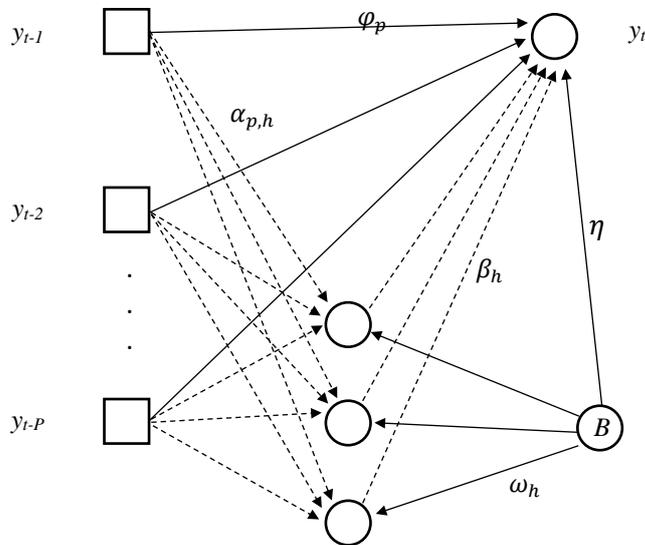
La estructura del modelo ARNN está presentada en la figura 1. Los parámetros del modelo: $\eta, \varphi_p, \beta_h, \omega_h, \alpha_{p,h}$ son estimados minimizando el error de regularización: λE_* , donde λ es un parámetro externo definido por el investigador; e_t vienen a ser los errores de pronóstico y_t^* y el valor que se observa y_t . T es el largo de la serie de tiempo y_t . Además,

$$E_* = |\eta| + \sum_{h=1}^H (|\beta_h| + |\omega_h|) + \sum_{p=1}^P |\varphi| + \sum_{p=1}^P \sum_{h=1}^H |\alpha_{p,h}|$$

El modelo y_t^* se reduce en un perceptrón multicapa imponiendo la restricción:

$\varphi_1 = \varphi_2 = \dots = \varphi_p$. La red neuronal se reduce a un modelo autorregresivo para un valor $H = 0$.

Figura 1
Modelo de una red neuronal artificial



Cuando se encuentran resultados contradictorios, surgen serios problemas al momento de definir el tipo de modelo más eficiente en la solución de problemas concretos de pronóstico.

Así, mientras algunos trabajos no encuentran diferencias entre los resultados hallados por unos y otros modelos, otros resultados tienden a apoyar una ligera superioridad de las redes neuronales sobre las técnicas estadísticas.

De este modo, se espera obtener resultados precisos para determinar qué modelo es más eficiente en la solución de problemas de pronóstico.

MATERIAL Y MÉTODOS

La población está constituida por los caudales máximos mensuales del río Huaura, desde su existencia hasta el mes de Julio del 2016.

La muestra está constituida por los caudales máximos mensuales del río Huaura, en el periodo hidrológico enero de 1960 a julio del 2016.

Para modelos ARIMA:

- El 90 % de los datos (enero 1960 - noviembre 2010) para modelar la serie de tiempo.
- El 10 % de los datos (diciembre 2010 - julio 2016) para pronóstico.

Para redes neuronales:

- El 60 % de los datos (enero 1960 - noviembre 1993) para entrenamiento.
- El 30 % de los datos (diciembre 1993 - noviembre 2010) para prueba.
- El 10% de los datos (diciembre 2010-Julio 2016) para pronóstico.

Marco muestral: Está constituido por los registros hidrológicos del río Huaura del Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI, 2016).

El procesamiento de datos se realizó mediante el uso de los siguientes softwares estadísticos:

- EViews 9: Para el modelamiento de la serie hidrológica por modelos ARIMA y cálculo de los pronósticos.
- IBM SPSS STATISTICS 24: Para el modelamiento de la serie hidrológica por redes neuronales artificiales autoregresivas, cálculos de pronósticos y comparación de pronósticos de ambos modelos con los valores reales.

RESULTADOS

Tabla 1

Estadísticas descriptivas de caudales máximos del río Huaura, enero de 1960 - julio de 2016

ESTADÍSTICO	VALOR
Media	26,53 m ³ /s
Mediana	26,36 m ³ /s
Máximo	76,54 m ³ /s
Mínimo	9,74 m ³ /s
Desv. Estandar	7,46 m ³ /s
Asimetría	1,36
Curtosis	6,70

Fuente: Elaboración propia

Se analizaron 679 registros mensuales de caudales máximos y se estimó el modelo (Tabla 2). Las estadísticas del comportamiento promedio de los caudales máximos del río Huaura es 26,53 m³/s, en febrero de 1974 llegó a un punto máximo con un valor de 76,54 m³/s; y en junio de 1960 se tuvo el punto mínimo con un valor de 9,74 m³/s, el comportamiento de los caudales máximos del río Huaura es asimétrica positiva a la derecha con un coeficiente de asimetría de 1,36 y es leptocúrtica porque su coeficiente de curtosis es mayor a 3 (Tabla 1).

Figura 2

Caudales máximos de río Huaura, enero de 1960 - noviembre de 2010

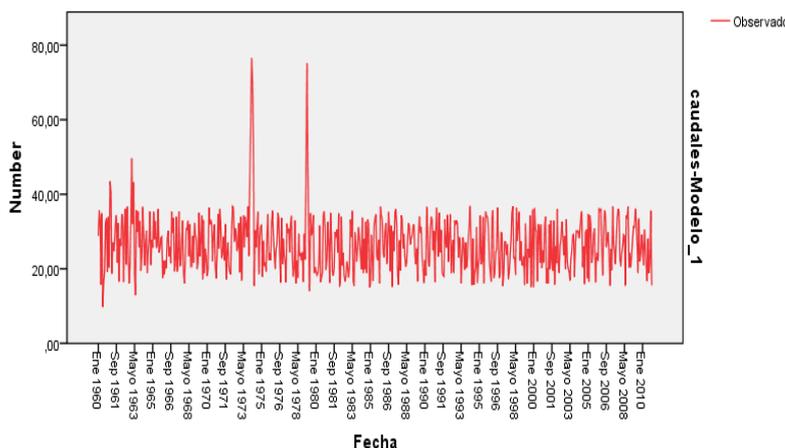


Tabla 2

Modelamiento de serie hidrológica del río Huaura mediante el modelo ARIMA

Variable	Coefficiente	Error Estándar	t-student	p-valor
Constante	26,62522	0,42994	61,92782	0,0000
AR(1)	0,225622	0,03963	5,693168	0,0000
MA(7)	0,121434	0,040518	2,997021	0,0028

R2: 62,2 % ; AIC: 6,82 ; Durbin-Watson: 1,99

Cuya ecuación es la siguiente:

$$\widehat{CMRH}_t = 26,6252 + 0,2256 CMRH_{t-1} + 0,1214 \varepsilon_{t-7}$$

Figura 3

Red Neuronal elegida para modelar y pronosticar los caudales máximos del río Huaura en el periodo enero de 1960 a julio de 2016

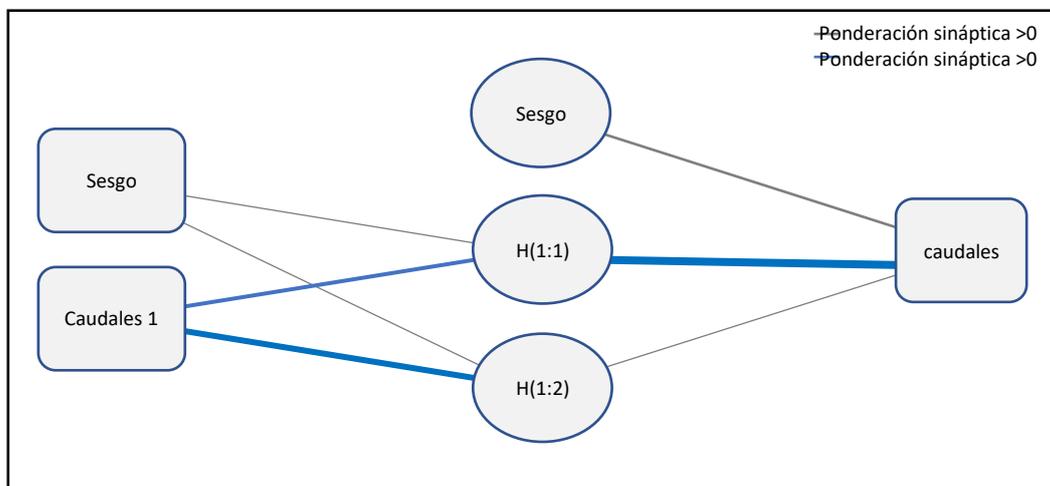


Tabla 3

Errores de pronóstico de los caudales máximos del río Huaura por modelos ARIMA y redes neuronales artificiales.

Modelo	DAM	EMC	PEMA	PME
ARIMA (1,0,7)	5,7	42,41	0,25	-0,1
Red Neuronal Artificial	5,68	42,28	0,24	-0,08

En la Tabla 3 se puede observar que el modelo más eficiente es el de Redes Neuronales al obtener errores de pronóstico menores en comparación a los modelos ARIMA. El pronóstico encontrado es insesgado debido a que los errores PEMA y PME son pequeños y próximos al cero.

DISCUSIÓN

La metodología de Box-Jenkins es empleada por Amarís et al (2017), en la aplicación de la técnica promedio móvil integrado autorregresivo, para analizar series hidrológicas anuales en el río Magdalena. Finalmente, pronosticó los caudales máximos y mínimos para el periodo del 2013 al 2024, oscilan entre 289,695 millones de m³ y 309,847 millones de m³, y entre 179,123 millones de m³ y 157,764 millones de m³ respectivamente, con una tendencia de decrecimiento de 106 millones de m³ en cien años. Esta simulación se comparó con los datos reales y mostraron un ajuste adecuado de los volúmenes mínimos y máximos. Si comparamos lo obtenido con nuestra investigación, podemos aseverar que la estimación con modelos ARIMA es precisa para periodos cortos, más no cuando la serie tiene gran volumen de datos.

Asimismo, en el trabajo de Chacón et al (2016), se empleó la técnica de redes neuronales para el pronóstico de caudales medios diarios en el río Amazonas, empleando el software Matlab, consiguiéndose una correlación de 0,97 y un mínimo error medio porcentual absoluto (MAPE) inferior al 0,10, en comparación al 0,24 de nuestra investigación. La modelación con esta técnica puede determinarse como buena gracias a capacidad de pronóstico comparado con otros métodos tradicionales del tipo lineal.

Finalmente, al comparar nuestra investigación con el trabajo de Rojo et al (2015), el cuál realizó el pronóstico de flujo de datos utilizando una previsión de sistemas combinados para obtener pronósticos individuales utilizando modelos ARIMA y redes neuronales artificiales autoregresivas, llegando a obtener cuadrados medios de error (CME) de 28,59 y 23,71 respectivamente; por lo tanto, al igual que nosotros se determinó que las redes neuronales realizan mejores pronósticos que los modelos ARIMA.

CONCLUSIONES

La serie hidrológica: Caudales máximos del río Huaura en el periodo de enero de 1960 a julio de 2016, pudo ser estimada empleando la metodología de Box-Jenkins con la técnica del promedio móvil integrado autorregresivo, obteniéndose la siguiente ecuación de estimación:

$$\widehat{CMRH} = 26,6252 + 0,2256 CMRH_{t-1} + 0,1214 \varepsilon_{t-7}$$

La Red Neuronal elegida para modelar y pronosticar los caudales máximos del río Huaura en el periodo enero de 1960 a julio de 2016, es un perceptrón multicapa el cual posee una capa de entrada, dos capas ocultas utilizando la tangente hiperbólica como función de activación y una capa de salida utilizando la función identidad.

Al comparar ambas técnicas se concluyó que, el modelamiento con redes neuronales es más eficiente que el modelo ARIMA al obtener menores errores de pronóstico.

RECOMENDACIONES

Se recomienda actualizar los modelos estimados periódicamente con la base de datos actualizada de la página web del SENAMHI.

Utilizar diferentes técnicas de pronóstico, donde se pueda evidenciar la minimización de la variabilidad de la serie para poder comparar y evaluar la eficiencia con las metodologías presentadas.

Las RNNa han generado un gran avance en la ciencia porque permiten realizar estimaciones y pronósticos más precisos, como es el caso de las series hidrológicas.

REFERENCIAS

- Amarís, G.; Ávila, H. & Guerrero, T. (2017). *Aplicación de modelo ARIMA para el análisis de series de volúmenes anuales en el río Magdalena*. *Revista Tecnura*, 21(52), 88-101.
- Catagora, E. (2008). *Predicciones Del Comportamiento de Caudales de la C.H. Machupicchu Mediante Análisis Arima de Series Temporales* (tesis de postgrado). Universidad Nacional de Ingeniería, Lima, Perú
- Caudales máximos del río Huaura. Recuperado el 20 de agosto de 2016 de <http://www.gob.pe/senamhi>
- Chacón, W., Valeriano, K., Ilachoque, J., & Sulla, J. (2016). Predicción de Caudales Medios Diarios en la Cuenca del Amazonas Aplicando Redes Neuronales Artificiales y el Modelo Neurodifuso ANFIS. *Research in Computing Science*, 113, 23-35.
- Fermín, W. (2010). Aplicación de Redes Neuronales Artificiales a la Modelización y Previsión de Caudales Medios Mensuales del río Huancané. *Revista Peruana GeoAtmosférica RPGA*, 2, 30-44.
- Gujarati, D., & Porter, D. (2010). *Econometria*. México: Mc-Graw Hill Interamericana.
- Hernández. R., Fernández .C., & Baptista. P. (2014). *Metodología de la investigación*. México, D.F.: McGraw-Hill Education.
- Laqui, W. (2010). Aplicación de Redes Neuronales Artificiales a la Modelización y Previsión de Caudales Medios Mensuales del Río Huancané. *Revista Peruana GeoAtmosférica RPGA* (2), 30-44
- Lujano, & Quispe, J. (2014). *Pronóstico de Caudales Medios Mensuales del Río Ilave Usando Modelos de Redes Neuronales Artificiales*. Universidad Nacional del Altiplano.
- Pita, E. (2008). *Pronóstico de Avenidas a Mediano Plazo Utilizando Redes Neuronales Artificiales y Modelos Tipo ARMA* (tesis de Maestría), Tecnológico de Monterrey, Campus Cuernavaca, Morelos.
- Previsión de Caudales Medios Mensuales del río Huancané. *Revista Peruana GeoAtmosférica RPGA*, 2, 30-44
- Rojo, J., Carbajal, L. & Velasquez, J. (2015). *Predicción de flujo de datos utilizando una previsión Sistema combinado* (Senior Member, IEEE), Universidad Nacional de Colombia, Sede Medellín, Colombia.
- Shumway, R., & Stoffer, D. (2011). *Time series analysis and its applications*. New York: Springer.
- Uriel, E., & Peiró, A. (2000). *Introducción al análisis de series temporales*. España: AC.
- Velásquez et al. (2011). ARNN: Un paquete para la predicción de series de tiempo usando redes neuronales autorregresivas. *Revista Avances en Sistema e Informática*, 2(8). Medellín, Colombia.
- Villacorta, S. Ochoa & M. Lauro, E. (2011). *Zonas Críticas por Peligros Geológicos en la Cuenca del Río Huaura*. Lima, Perú: INGEMMET.