

Predicción de caudales medios diarios en la cuenca del Amazonas aplicando redes neuronales artificiales y el modelo neurodifuso ANFIS

Walter Enrique Béjar Chacón, Kid Yonatan Valeriano Valdez,
Julio Cesar Ilachoque Umasi, Jose Sulla Torres

Universidad Nacional San Agustín, Escuela Profesional de Ingeniería de Sistemas,
Arequipa, Perú

{wbejarch, jim.kidv, jcilachoque, josullato}@gmail.com

Resumen. El presente artículo muestra los resultados obtenidos al aplicar redes neuronales y el modelo neurodifuso ANFIS para la predicción de caudales medios diarios en una sección de la cuenca del Amazonas. Se aplicaron los pasos que componen la metodología KDD - Knowledge Discovery in Databases sobre la información hidrológica recolectada de las estaciones del servicio de observación SO HYBAM y sobre la información climatológica recolectada del Instituto de Investigación Internacional de Clima y Sociedad. La simulación se llevó a cabo en el software Matlab® donde se evaluó el comportamiento de ambos modelos, el estudio mostró que al utilizar técnicas de Inteligencia Artificial se consiguió coeficientes de correlación - CC superiores al 97%, un error medio porcentual absoluto - MAPE por debajo del 10% y otras 4 métricas que fueron utilizadas en esta investigación. Los resultados muestran que estos modelos pueden etiquetarse como buenos modelos para la predicción, gracias a su capacidad predictiva en comparación con métodos tradicionales del tipo lineal.

Palabras clave: Predicción de caudales, redes neuronales artificiales, sistemas neurodifusos, ANFIS, inteligencia artificial.

Forecast of Average Daily Flows in the Amazon Basin Using Artificial Neural Networks and the Adaptive Neuro-fuzzy Inference System (ANFIS)

Abstract. This article shows the results obtained by applying neural networks and neurofuzzy model ANFIS for predicting average daily flows in a section of the Amazon basin. The steps in the KDD methodology (Knowledge Discovery in Databases) were applied on hydrological data collected from observation stations of the SO HYBAM and climatological information collected from the International Research Institute for Climate and Society. The simulation was performed in the software Matlab® where the behavior of both models were evaluated, the study showed that using artificial intelligence techniques achieved correlation coefficients - CC above 97%, a mean absolute percentage error -

MAPE by below 10% and 4 other metrics that were used in this research. The results show that these models can be labeled as good models for prediction, thanks to its predictive ability compared to traditional methods of linear type.

Keywords: Forecast prediction, artificial neural networks, fuzzy neuro systems, ANFIS, Artificial Intelligence.

1. Introducción

Diversos trabajos de investigación tales como [1, 2] muestran que la cuenca Amazónica ha sufrido durante las dos últimas décadas de estiajes fuertes pero también de grandes inundaciones (1999, 2009, 2012, 2014), cada uno de estos eventos fue catastrófico para las poblaciones ribereñas de los grandes ríos Amazónicos, en el Perú los niveles alcanzados en el año 2015 por el Río Amazonas llegaron a equivalentes de crecidas históricas.

Lo que lleva consigo a aplicar métodos de pronóstico hidrológicos generalmente usando regresiones lineales, dichos modelos miden la relación entre variables dependientes e independientes del fenómeno y utilizan los caudales como datos de entrada [3, 4]. Pero debido a la no linealidad de estos fenómenos y que adicionalmente para que estos modelos sean más precisos se necesita incluir otras variables del tipo hidrológico y morfológico, lo que hace que estos modelos no sean apropiados [5] y sea necesario utilizar otro tipo de modelos.

Actualmente, se han desarrollado diversos estudios de modelos de predicción que integran técnicas de Inteligencia Artificial, cuya estructura matemática flexible es capaz de identificar relaciones complejas no lineales entre los datos de entrada y salida, y en problemas en donde es difícil describir el proceso utilizando ecuaciones físicas [6]. Una de las técnicas más usadas en este campo son las redes neuronales artificiales - ANN que simulan el funcionamiento del cerebro para la resolución de problemas, otras técnicas de Soft Computing utilizadas son la lógica difusa, que permite tratar información imprecisa y también el modelo neurodifuso ANFIS, que es una combinación de la lógica difusa y las redes neuronales [7, 8].

Recientes estudios han desarrollado distintos modelos implementando técnicas de Inteligencia Artificial, como la realizada en Grecia en el Observatorio Nacional de Atenas (NOA), donde se pronosticó la precipitación máxima diaria aplicando modelos de redes neuronales artificiales, utilizando datos durante los años de 1891-2009. Los resultados de la investigación sugieren que dependiendo de la frecuencia del entrenamiento puede tener un impacto en la formación óptima de la predicción [9].

Partiendo de este panorama se presenta esta investigación desarrollada en una sección de la cuenca del Amazonas, cuyo objetivo es el pronóstico de caudales medios diarios aplicando redes neuronales artificiales y el modelo neurodifuso ANFIS, con los cuales se podría realizar estudios para el pronóstico de inundaciones.

2. Aspectos metodológicos

El proceso que siguió esta investigación fue el conocido Knowledge Discovery in Databases - KDD, el cual pone un especial énfasis en la búsqueda de patrones

comprensibles que pueden ser interpretados como conocimiento útil o interesante. Se siguieron los pasos que componen el proceso KDD, que se pueden observar en la Fig. 1 y que formaron la base para la obtención de resultados en esta investigación.

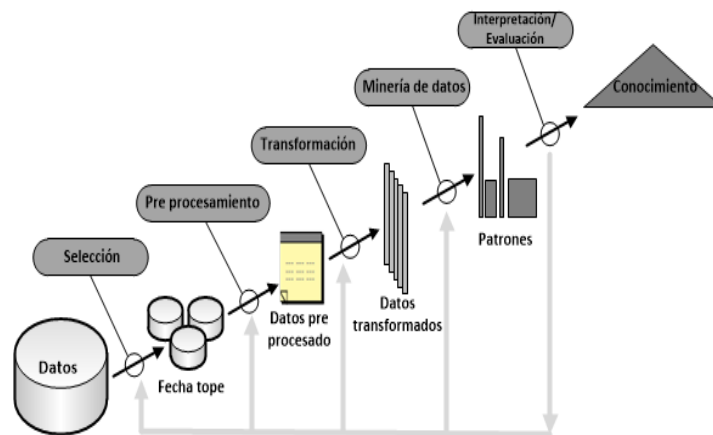


Fig. 1. Pasos que componen el proceso KDD [10].

2.1 Información hidrológica y climatológica

La cuenca hidrológica del Río Amazonas se ubica sobre terrenos de varios países de América del Sur: Perú, Colombia, Bolivia, Ecuador, Venezuela, Guyana, Surinam y Brasil, este último con la mayor extensión de cerca de 4 millones de km², que corresponden a casi la mitad de su superficie, la cuenca cubre una superficie de 6,2 millones de km². Así, la cuenca del Amazonas es la mayor cuenca hidrográfica en el mundo con un volumen de agua impresionante.

Se realizó una búsqueda de fuentes de datos hidrológicos libres, gracias al Servicio de Observación SO HYBAM (anteriormente Observatorio de Investigación del Medio ambiente) se pudo tener acceso a información de caudales de más de veintiuna estaciones que miden el caudal de la cuenca del Río Amazonas.

En cuanto a información climatológica, esta pudo ser obtenida gracias al Instituto Internacional de Investigación sobre el Clima y Sociedad, este posee un gran conjunto de datos muchos de ellos relacionados a la climatología, para esta investigación se utilizó el conjunto de datos brindado por la NOAA - National Oceanic and Atmospheric Administration, conjunto de datos que entre los atributos que posee se encuentra el atributo de precipitación media diaria [11].

2.2 Selección de datos

Para el presente estudio se escogió los datos de caudales medios diarios de 4 estaciones, las cuales son: Obidos, Fazenda Vista Alegre, Manacapuru y Serrinha, cuya localización se puede observar en la Fig. 2.

Dichas estaciones fueron escogidas ya que presentaban el menor número de datos faltantes respecto a los intervalos de tiempo en común.

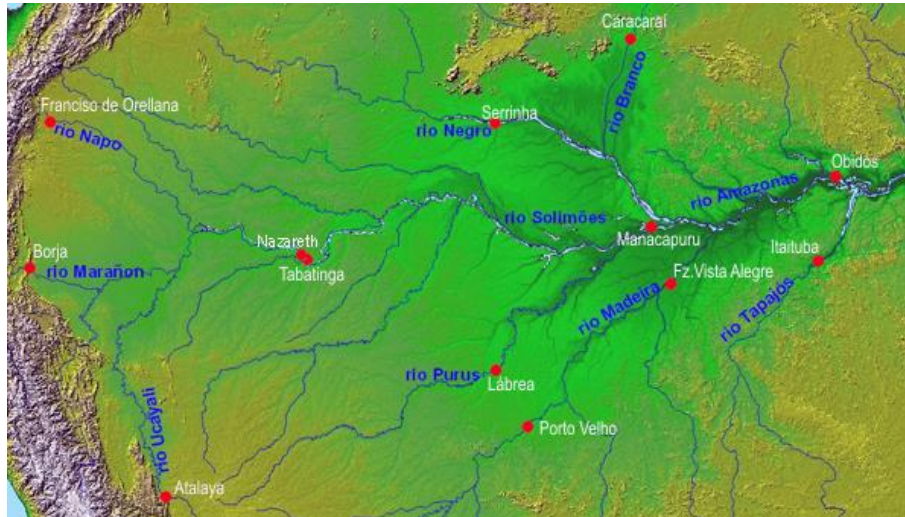


Fig. 2. Estaciones SO HYBAM ubicadas en la Cuenca del Amazonas.

En cuanto a los datos de precipitación solo se tomó en cuenta los datos obtenidos de una estación que está ubicada muy cerca de la desembocadura del Río Branco, dichos datos están comprendidos desde Marzo de 1989 a Diciembre de 1999, por lo que los datos de caudal fueron restringidos a dichas fechas.

En la Fig. 3 podemos ver los datos hidrológicos de la estación Obidos, dichos datos serán usados como salida de los modelos antes mencionados, ya que la porción del río donde se ubica esta estación es alimentada por el caudal de las otras 3 estaciones.

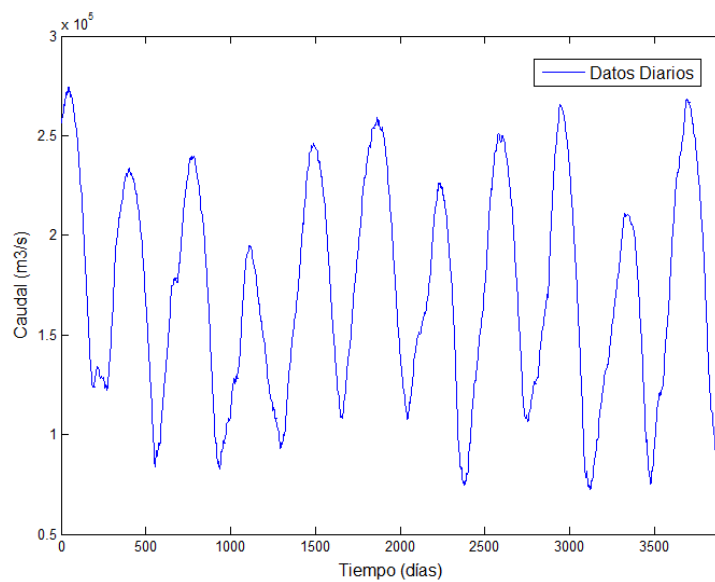


Fig. 3. Hidrograma estación Obidos 1989-1999.

2.3 Preprocesamiento

Como se pudo apreciar en la Fig. 3 los datos no presentan ruidos por lo que no es necesario realizar ninguna técnica de suavizamiento para los datos.

El inconveniente que se presentó en dicha etapa fue que había intervalos de tiempo en los cuales no se había registrado el caudal medio diario, esto se pudo dar por diferentes razones, como falla en los equipos de medición, falla en los registros de datos, etc. La cantidad de registros faltantes no sobrepasaba el 5% de la cantidad total de registros, por lo que los datos todavía podían ser utilizados [12], ya que la cantidad de registros faltantes no era significativa se procedió a estimar dichos valores en base a periodos de tiempo similares.

2.4 Transformación

Para esta etapa se optó por normalizar los datos, que consiste en re-escalar los valores de los datos dentro de un rango específico tal como -1 a 1 o de 0 a 1 [12][13].

Las ventajas de realizar una normalización a los datos hace posible acelerar la etapa de aprendizaje y evitar problemas numéricos tales como pérdida de precisión y desbordamiento aritméticos (overflows).

$$V' = 1/(1 + e^{-a}) \quad (1)$$

$$a = (V - \text{mean})/\text{std} \quad (2)$$

Para los datos seleccionados se decidió realizar la normalización SoftMax que nos devuelve un rango de salida de 0 a 1. La normalización SoftMax utiliza las ecuaciones (1) y (2) para normalizar los datos de entrada, donde V : dato a normalizar, mean : media de los datos a normalizar, std : desviación estándar de los datos a normalizar y V' : dato normalizado.

2.5 Modelos de predicción

A continuación se describen las diferentes herramientas de ajuste y las metodologías de predicción que se usaron en esta investigación.

2.5.1 Redes neuronales artificiales

Las redes neuronales artificiales son modelos matemáticos inspirados en procesos neurobiológicos en los que el análisis de la información se imita a las acciones desarrolladas por las neuronas en el cerebro. La estructura estándar de una red neuronal artificial está compuesta por un conjunto de neuronas organizadas en capas (entrada, ocultas y de salida), distribuidas jerárquicamente; constituyendo un sistema funcional autónomo.

En este sistema inteligente, como el mostrado en la Fig. 4; se identifican los siguientes elementos: variables de entrada y salida, pesos sinápticos (que son la intensidad de interacción entre las neuronas), función de propagación, función de activación y función de salida. Además, se debe tener en cuenta el número de capas y

neuronas, ya que estos son los parámetros más importantes en la red neuronal artificial determinando así la eficiencia del sistema.

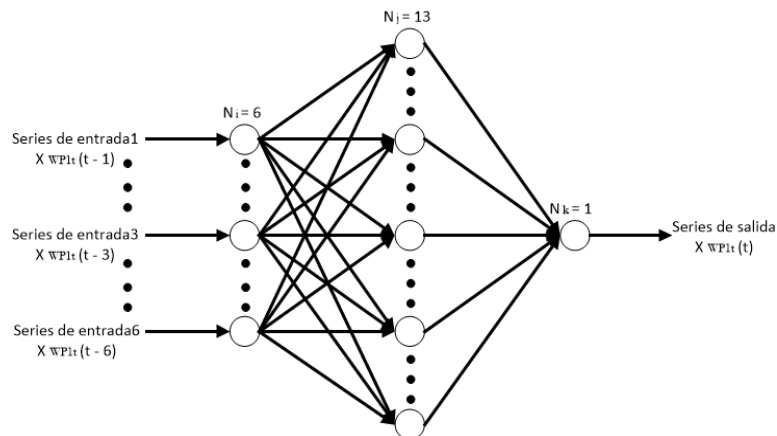


Fig. 4. Red Neuronal de 3 Capas [5].

Una de las ventajas de las redes neuronales artificiales es que pueden ser una herramienta útil para el modelado, cuando la relación entre los datos de entrada y salida no es clara o en su defecto es desconocida, razón por la cual este tipo de modelos son llamados de caja negra; lo que nos permite que a través de sus composiciones matemáticas sean capaces de modelar sistemas complejos como los sistemas hidrológicos. Otro de los beneficios derivados de las redes neuronales artificiales es su capacidad de generar salidas de una combinación específica de entradas y su capacidad de respuesta frente al manejo de datos no lineales, por lo que estos sistemas inteligentes pueden llegar a ser mejores que los sistemas de modelos lineales debido a su flexibilidad al abordar problemas complejos.

La simulación de las redes neuronales artificiales se realizó en el toolbox del software Matlab 2013® llamado nntool (neural network tool). La estructura de la red neuronal se compone de 4 conjuntos de datos de entrada (3 conjuntos de datos de caudal y un conjunto de datos de precipitación) y un conjunto de datos de salida. En cuanto a los parámetros utilizados para la calibración del modelo se estableció un tipo de entrenamiento backpropagation [14], debido a que la arquitectura implementada fue multi-capa, con neuronas ocultas, se utilizó la función de aprendizaje Levenberg-Marquardt (*trainlm*) que realiza mejor la función de ajuste para el reconocimiento de patrones del sistema y la función de error medio cuadrático (MSE).

Se utilizó un máximo de mil repeticiones para correr el modelo hasta llegar a la validación total del procesamiento de la información en conjunto con un gradiente mínimo de $1e - 07$ y un máximo de seis revisiones de validación para evaluar la calidad del modelo. El modelo computacional divide los datos en tres muestras, el 70% del total de los datos ingresados sirven para el entrenamiento, 15% de los datos para la validación y el otro 15% para las pruebas del modelo.

Se procedió a simular distintos escenarios en los cuales se modificaba el número de capas (entre 2 y 5) y el número de neuronas (entre 3 y 10), generando así treinta y dos escenarios en los cuales se utilizó la función de propagación *sigmoidal - sigmoidal*.

En cada escenario entrenado se realizó la simulación del pronóstico durante el rango de fechas elegido y con los resultados obtenidos se hallaron cada una de las métricas de comparación por medio de un programa realizado en Matlab 2013®.

2.5.2 Modelo neurodifuso ANFIS

Para explicar que es un sistema neuro-difuso se debe partir del enfoque de la lógica difusa que se basa en expresiones lingüísticas inciertas en lugar de la incertidumbre numérica. Esta técnica se basa en un sistema de inferencia difusa (FIS) basado en tres componentes: una base de normas (rulebase) que contiene las reglas difusas si-entonces, una base de datos (database) definida por una función de pertenencia y un sistema de inferencia (interference system) que combina las reglas difusas y produce los resultados del sistema [15], la estructura de dicho sistema se muestra en la Fig. 5.

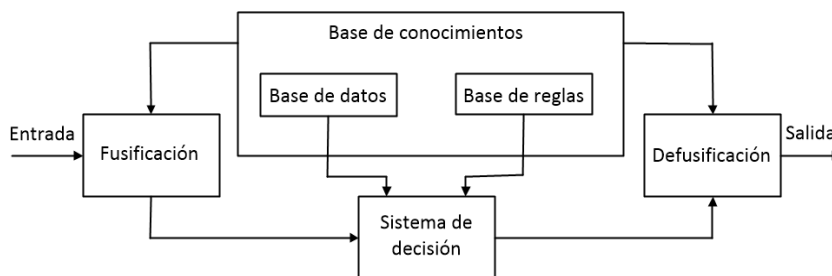


Fig. 5. Estructura Modelo de Inferencia Difuso [15].

La estructura de un sistema neuro-difuso consiste en la unión de dos técnicas artificiales inteligentes: redes neuronales artificiales y lógica difusa, razón por la cual es llamada ANFIS (Adaptive Neuro-Fuzzy Interference System). Este sistema utiliza los algoritmos de aprendizaje y la configuración por medio de capas y neuronas de las redes neuronales artificiales, mientras que de los sistemas de inferencia difusa utiliza el razonamiento difuso que permite generar reglas de inferencia a partir de la asignación de variables lingüísticas a la información caracterizando así la combinación de cada una de las entradas a una o varias salidas. Esta combinación permite a la red organizarse a sí misma y generar adaptabilidad del sistema difuso para resolver distintos problemas, en la Fig. 6 se muestra la estructura general del modelo ANFIS que se compone de 5 capas.

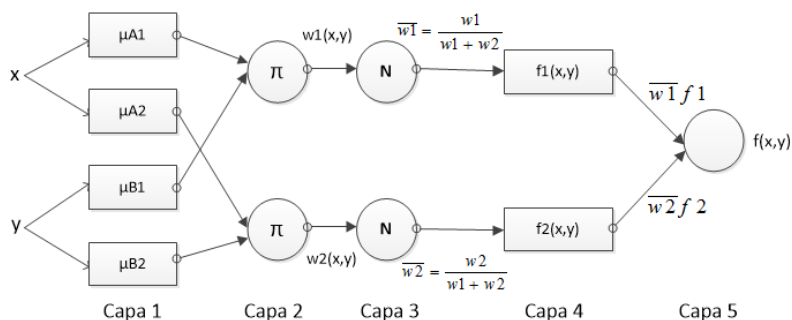


Fig. 6. Estructura Modelo ANFIS [16].

La simulación del modelo ANFIS se realizó en el toolbox del software Matlab 2013® llamado anfisedit. La estructura del modelo ANFIS se compone de 4 conjuntos de datos de entrada (3 conjuntos de datos de caudal y un conjunto de datos de precipitación) y un conjunto de datos de salida.

El conjunto de datos de entrada se dividió de la siguiente forma, el 70% del total de los datos ingresados sirven para el entrenamiento, 15% de los datos para la validación y el otro 15% para las pruebas del modelo. Se simularon en total 18 escenarios, variando la cantidad de épocas entre 10 y 15.

Respecto a las reglas de inferencia se usó un sistema tipo Sugeno, usando reglas de inferencia tipo producto (and) entre las entradas, funciones de pertenencia trapezoidal, campana y triangular junto a un método de defusificación tipo *wtaver* el cual retorna la media ponderada a la salida del sistema difuso del ANFIS.

3. Resultados

Para comparar la precisión y exactitud de los modelos inteligentes mostrados en esta investigación se usaron 6 datos estadísticos, usados en la mayoría de investigaciones observados y utilizados como métricas de evaluación de modelos simulados [5]. Estos son:

MAE (Mean Absolute Error): el error medio absoluto es una medida de precisión usada para evaluar pronósticos, en él se evalúa el valor absoluto promedio de la diferencia entre el dato real y el dato pronosticado, entre más pequeño sea el error o tienda a cero, será más preciso el pronóstico [17].

MSE (Mean Squared Error): el error medio cuadrático es una medida de precisión usada para evaluar pronósticos, en la que se evalúa el desempeño del valor promedio de la diferencia entre el dato real y el dato pronosticado al cuadrado [18].

MAPE (Mean Absolute Percentage Error): el error porcentual absoluto de la media permite analizar la exactitud del modelo en términos porcentuales, teniendo en cuenta el valor absoluto de la relación entre el dato real y el dato pronosticado. La escala para evaluar la exactitud del modelo usando el MAPE determina que un pronóstico muy exacto es el que tiene un valor menor o igual al 10%, un buen pronóstico tiene un valor entre un rango del 11% al 20%, un pronóstico razonable entre el 21% al 50% y un pronóstico inadecuado mayor al 50% [19].

RMSE (Root Mean Squared Error): la raíz cuadrada del error cuadrático medio es una medida de precisión usada para evaluar pronósticos, en ella se evalúa el valor de la raíz del promedio cuadrático de la diferencia entre el dato real y el dato pronosticado, entre más pequeño sea el error o tienda a cero, será catalogado como el pronóstico más preciso [17].

CC (Correlation Coefficient): el coeficiente de correlación de Pearson determina la relación del dato real y el dato pronosticado, en cuanto a la covarianza con las desviaciones de los dos tipos de datos, a diferencia del CCC, el coeficiente de correlación ignora componentes de exactitud [20].

CCC (Concordance Correlation Coefficient): el coeficiente de correlación de concordancia indica la relación entre la precisión y exactitud del modelo, este mide el grado en que la covarianza del dato real y el dato pronosticado se acercan a la recta de 45° del modelo, este factor se evalúa con valores entre 0 y 1 [21].

3.1 Resultados redes neuronales artificiales

Se eligieron 5 de los 32 escenarios con mayor CCC, dichos escenarios se compararon entre sí ponderando aquellas métricas que cumplían con la mayor cantidad de resultados favorables en los criterios estadísticos a evaluar con el MAE, MAPE, MSE, RMSE y CC para calcular la precisión y exactitud de los modelos. A continuación en la Tabla 1 se observan los 5 mejores escenarios.

Tabla 1. Mejores escenarios redes neuronales artificiales.

Escenario	CCC	CC	MAE	MAPE	MSE	RMSE
Esc 32	0.9937	0.9937	0.0173	4.5384	0.00061	0.0246
Esc 27	0.9934	0.9934	0.0182	4.8242	0.00064	0.0252
Esc 28	0.9932	0.9933	0.0182	5.0245	0.00065	0.0254
Esc 24	0.9930	0.9930	0.0183	4.8438	0.00067	0.0259
Esc 20	0.9920	0.9921	0.0198	5.1768	0.00076	0.0276

Luego de analizar todos los criterios en los 5 mejores escenarios, el mejor escenario es el número 32 debido a que cumple con la mayor cantidad de resultados favorables, menor MAE (0.0173), menor MAPE (4.5384%), menor MSE (0.00061), menor RMSE (0.0246), mayor CC (0.9937) y mayor CCC (0.9937). Tomando como referencia el resultado obtenido con el MAPE se puede decir que el pronóstico arrojado es exacto ya que se encuentra por debajo del 10%, el CC nos indica que el modelo presenta el 99% de precisión del pronóstico en cuanto a la relación de los datos reales con los datos simulados. En este escenario se implementaron 5 capas y 10 neuronas por cada capa, lo cual nos indica que al implementar una red neuronal artificial se debe evaluar el incrementar el número de capas y de neuronas para un pronóstico adecuado.

A continuación se observa en la Fig. 7 donde se muestra la comparación entre los datos reales con el mejor escenario simulado. En este histograma se muestran los caudales medios diarios en el periodo comprendido desde Febrero de 1996 a Diciembre de 1999, que corresponde al 15% de datos de prueba. Se observa que el modelo basado en redes neuronales artificiales se acerca bastante a los datos de caudal real.

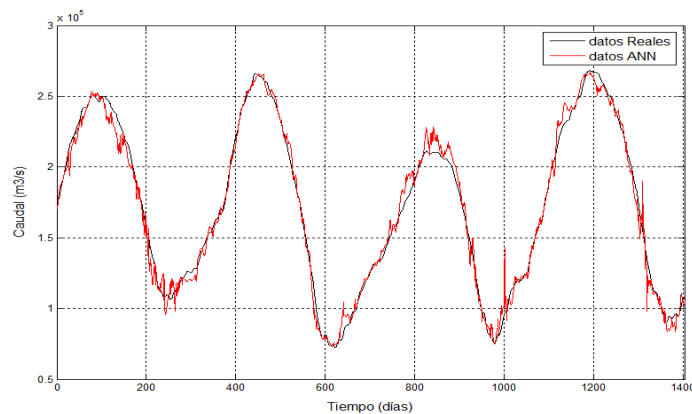


Fig. 7. Comparación Hidrogramas Reales vs ANN.

3.2 Resultados modelo neurodifuso ANFIS

De igual manera se escogieron los 5 mejores escenarios con el sistema neurodifuso, que son mostrados en la Tabla 2.

Tabla 2. Mejores escenarios modelo neurodifuso.

Escenario	CCC	CC	MAE	MAPE	MSE	RMSE
Esc 18	0.977	0.9772	0.0366	9.2452	0.0022	0.0465
Esc 17	0.9769	0.9772	0.0367	9.2607	0.0022	0.0466
Esc 16	0.9767	0.977	0.0368	9.3137	0.0022	0.0468
Esc 15	0.9763	0.9766	0.0371	9.3736	0.0022	0.0471
Esc 5	0.9762	0.9765	0.0373	9.3855	0.0022	0.0473

Luego de observar los resultados de la Tabla 3, se eligió el escenario 18 como el mejor pronóstico ya que cumple con la mayor cantidad de resultados favorables, menor MAE (0.0366), menor MAPE (9.2452%), menor MSE (0.0022), menor RMSE (0.0465), mayor CC (0.9772) y el mayor CCC (0.977). Este escenario fue simulado con una función de pertenencia de campana, con 15 épocas. La comparación con el mejor de los escenarios se muestra en la Fig. 8 donde figuran los caudales reales comprendidos en el periodo que va desde Febrero de 1996 a Diciembre de 1999, dichos datos corresponden al 15% de datos de prueba.

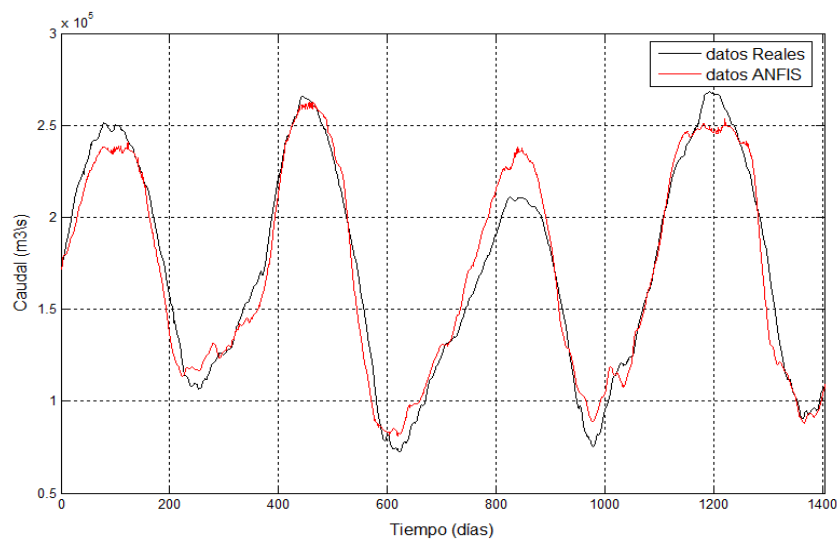


Fig. 8. Comparación Hidrogramas Reales vs ANFIS.

3.3 Resultados ANFIS vs red neuronal artificial

A continuación se muestra en la Tabla 3 los dos mejores modelos que se desarrollaron en esta investigación (red neuronal artificial sigmoidal - sigmoidal, modelo ANFIS con función de pertenencia en campana).

Tabla 3. Comparación del mejor escenario de los modelos planteados.

Modelo	CCC	CC	MAE	MAPE	MSE
Modelo ANN	0.9937	0.9937	0.0173	4.5384	0.00061
Modelo ANFIS	0.977	0.9772	0.0366	9.2452	0.0022

Los resultados obtenidos muestran una clara diferencia entre el modelo neurodifuso y el modelo basado en redes neuronales artificiales, siendo este último el que logra entregar los mejores pronósticos. Para observar mejor el comportamiento de los datos y el porqué de la elección del CCC, es necesario graficar los datos reales vs datos simulados comparándolos con la recta de la ecuación $y=x$.

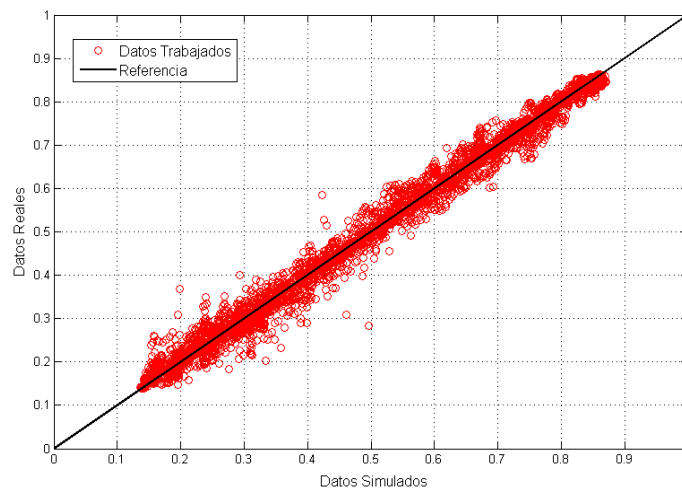


Fig. 9. Datos Reales vs Datos Simulados ANN.

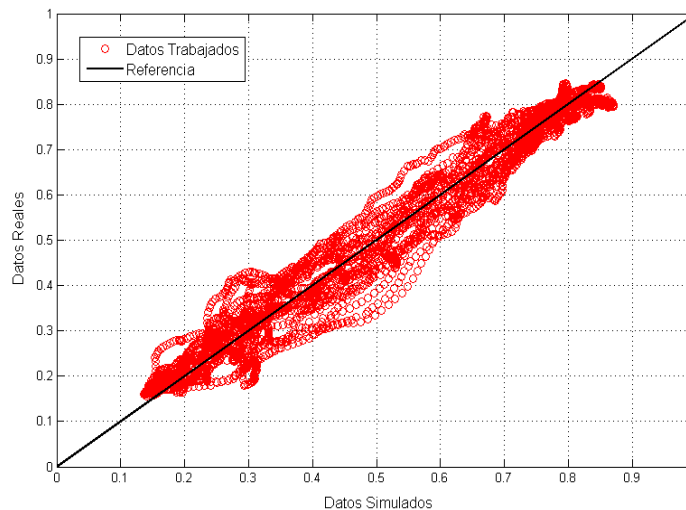


Fig. 10. Datos Reales vs Datos Simulados ANFIS.

En la Fig. 9 se hace la comparación de los datos reales y simulados por la red neuronal artificial donde podemos observar que los datos no están tan dispersos en comparación a la Fig. 10, en donde los datos se alejan de la recta de referencia. Gracias a las métricas utilizadas podemos observar que entre el dato pronosticado y el dato real el error es inferior al 10%.

4. Conclusiones

La implementación de modelos basados en técnicas de inteligencia artificial con los cuales se simuló el comportamiento de un tramo de la cuenca del Amazonas arrojó resultados estadísticos que demuestran pronósticos muy cercanos a los datos reales y la efectividad de dichas técnicas (MAPE entre 4% y 10% que indica un buen pronóstico; CC entre 0.97 y 0.99 que indica una buena relación entre los datos reales y los simulados; CCC entre 0.97 y 0.99 que indica demuestra precisión y exactitud en el pronóstico).

Se logró observar que las técnicas de inteligencia artificial (o de caja negra) como las redes neuronales artificiales arrojaron mejores resultados que los modelos basados en ANFIS, los modelos basados en redes neuronales artificiales pueden ser iguales o mejores que los modelos hidrológicos comunes basados en modelos no lineales y que la aplicación de este tipo de soluciones puede ser una buena alternativa para la generación de sistemas de alertas tempranas, integrando dicho sistema a sistemas de información geográfica que permitan visualizar la información pronosticada en estos sistemas.

Se recomienda continuar la investigación para la construcción de alertas tempranas de sistemas de información geográfica (SIG) y buscar otras zonas donde se pueda adaptar este proyecto para la preparación de la población ante riesgos de inundación.

Además, se recomienda la implementación del uso de técnicas artificiales inteligentes enfocadas al pronóstico de inundaciones en países de relieves tan diversos como los del continente Suramericano, es necesario continuar validando estos métodos no solo con los datos de 3 estaciones hidrológicas si no también integrando las medidas de estaciones cercanas o tener en cuenta otro tipo de variables tales como sociales, logísticas, económicas y culturales, lo que podría generar mayores desarrollos tecnológicos para no solo la creación de alertas tempranas sino también para la prevención de desastres naturales de distinto tipo.

Por otro lado surge la posibilidad de estudiar este problema con otro tipo de técnicas, lógica difusa intuitiva y algoritmos genéticos, comparando estos con modelos matemáticos y/o hidráulicos.

Referencias

1. Espinoza, J.C., Ronchail, J., Frappart, F., Lavado, W., Santini, W., Guyot, J.L.: The major floods in the amazonas river and tributaries (western amazon basin) during the 1970-2012 period: A focus on the 2012 flood*. *Journal of Hydrometeorology*, Vol. 14, No. 3, pp.1000–1008 (2013)
2. Marengo, J.A., Espinoza, J.C.: Extreme seasonal droughts and floods in amazonia: causes, trends and impacts. *International Journal of Climatology* (2015)

3. Pandey, G.R., Nguyen, V.T.V.: A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, Vol. 225, No. 1, pp. 92–101 (1999)
4. Weisberg, S.: *Applied linear regression*. John Wiley & Sons, Vol. 528 (2005)
5. Dawson, C.W., Abrahart, R.J., Shamseldin, A.Y., Wilby, R.L.: Flood estimation at ungauged sites using artificial neural networks. *Journal of Hydrology*, Vol. 319, No. 1, pp. 391–409 (2006)
6. Seckin, N., Cobaner, M., Yurtal, R., Haktanir, T.: Comparison of artificial neural network methods with l-moments for estimating flood flow at ungauged sites: the case of east mediterranean river basin, turkey. *Water resources management*, Vol. 27, No. 7, pp. 2103–2124 (2013)
7. Kalteh, A.M.: Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Computers & Geosciences*, Vol. 54, pp. 1–8 (2013)
8. Goyal, M.K., Bharti, B., Quilty, J., Adamowski, J., Pandey, A.: Modeling of daily pan evaporation in sub tropical climates using ann, ls-svr, fuzzy logic, and anfis. *Expert Systems with Applications*, Vol. 41, No.11, pp. 5267–5276 (2014)
9. Nastos, P.T., Paliatatos, A.G., Koukoultsos, K.V. Larissi, I.K., Moustris, K.P.: Artificial neural networks modeling for forecasting the maximum daily total precipitation at athens, greece. *Atmospheric Research*, Vol. 144, pp. 141–150, (2014)
10. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI magazine*, Vol. 17, No. 3, p. 37 (1996)
11. Vose, R.S., Schmoyer, R.L., Peterson, T.C., Steurer, P.M., Heim Jr, R.R., Karl, T.R., Eischeid, J.K.: *The global historical climatology network: Long-term monthly temperature, precipitation, and pressure data*. Technical report, Oak Ridge National Lab., TN (United States) (1992)
12. Hernández, C., Rodríguez Rodríguez, J.E.: Preprocesamiento de datos estructurados. *Vínculos*, Vol. 4, No. 2, pp. 27–48 (2013)
13. Dapozo, G.N., Porcel, E., López, M.V., Bogado, V.S.: Técnicas de preprocesamiento para mejorar la calidad de los datos en un estudio de caracterización de ingresantes universitarios. *IX Workshop de Investigadores en Ciencias de la Computación* (2007)
14. Kourentzes, N., Barrow, D.K., Crone, S.F.: Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, Vol. 41, No. 9, pp. 4235–4244 (2014)
15. Firat, M., Güngör, M.: River flow estimation using adaptive neuro fuzzy inference system. *Mathematics and Computers in Simulation*, Vol. 75, No. 3, pp. 87–96, (2007)
16. Firat, M.: Comparison of artificial intelligence techniques for river flow forecasting. *Hydrology and Earth System Sciences*, Vol. 12, No. 1, pp. 123–139 (2008)
17. Singhal, D., Swarup, K.S.: Electricity price forecasting using artificial neural networks. *International Journal of Electrical Power & Energy Systems*, Vol. 33, No. 3, pp. 550–555 (2011)
18. S. da S. Gomes, G., Ludermir, T.B.: Optimization of the weights and asymmetric activation function family of neural network for time series forecasting. *Expert Systems with Applications*, Vol. 40, No. 16, pp. 6438–6446 (2013)
19. Lewis, C.D.: *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth Heinemann (1982)
20. Li, J., Feng, P., Wei, Z.: Incorporating the data of different watersheds to estimate the effects of land use change on flood peak and volume using multi-linear regression. *Mitigation and Adaptation Strategies for Global Change*, Vol. 18, No. 8, pp. 1183–1196 (2013)
21. Lin, L., Hedayat, A.S., Sinha, B., Yang, M.: Statistical methods in assessing agreement: Models, issues, and tools. *Journal of the American Statistical Association*, Vol. 97, No. 457, pp. 257–270 (2002)