

Development of a Semi-distributed Hydrological Model at daily level using Neural Networks

Carlos Mendoza, Ing. Civil¹, Juan Cabrera, Magiste²

^{1,2}Universidad Nacional de Ingeniería, Perú, cjmmendoza@hotmail.com, juancabrera@uni.edu.pe

Abstract— A semi-distributed daily precipitation-runoff model was developed and applied to the sub-basins formed by the Barrios, Puente Paltashaco, Teodulo Peña, Malvados and Obrajillo hydrometric stations. For the development, use was made of neural networks of the Pre-fed and NARX (nonlinear autoregressive network with exogenous inputs) type in their open (Open NARX) or parallel series and closed (Closed NARX) or parallel forms, where it has been observed that the NARX network, specifically its parallel series version, has obtained the best results in the performance indicators, these being the Nash Sutcliffe Efficiency and the Peak Difference.

Keywords—Artificial Neural Network, hydrology, hydrological model.

Desarrollo de un Modelo Hidrológico Semidistribuido a nivel diario mediante Redes Neuronales

Carlos Mendoza, Ing. Civil¹, Juan Cabrera, Magiste²

^{1,2}Universidad Nacional de Ingeniería, Perú, cjmendoza@hotmail.com, juanabrera@uni.edu.pe

Resumen— Se ha desarrollado un modelo semidistribuido de precipitación escurrentía a nivel diario el cual ha sido aplicado a las subcuencas formadas por las estaciones hidrométricas Barrios, Puente Paltashaco, Teodulo Peña, Malvados y Obrajillo. Para el desarrollo se hizo uso de las redes neuronales del tipo Preatimentada y NARX (siglas de “nonlinear autoregressive network with exogenous inputs”) en sus formas abierta (Open NARX) o de series en paralelo, y cerrada (Closed NARX) o en paralelo, donde se ha observado que la red NARX, específicamente su versión de series paralelas ha obtenido los mejores resultados en los indicadores de desempeño, siendo estos la Eficiencia de NashSutcliffe y la Diferencia de Picos.

Palabras clave—Redes neuronales, hidrología, modelo hidrológico.

I. INTRODUCCIÓN

En el marco de la realidad peruana, donde existe una baja densidad de estaciones meteorológicas en muchas áreas del país, sobre todo en estaciones hidrométricas, la variedad de modelos hidrológicos que puedan ser aplicados resulta limitada. Por otra parte, estos modelos son en su mayoría desarrollados en el extranjero, en condiciones no necesariamente semejantes a las particularidades de esta región, ocasionando así una mayor incertidumbre en los resultados y, en consecuencia, pérdida de confiabilidad en los mismos, afectando de esta manera los diseños de infraestructura hidráulica, los cuales dependen de las salidas de estos modelos hidrológicos.

Así mismo, otro factor limitante en la simulación de caudales es la falta de información relativa a la cobertura, tipos de suelo, infiltración u otras variables asociadas, las cuales son requeridas por modelos más complejos como los modelos distribuidos y de base física. Al respecto, el uso de redes neuronales artificiales surge como una alternativa de bajo requerimiento de datos bajo un esquema de “caja negra”. Durante los últimos años se ha ido acrecentando el número de estudios referente a la aplicabilidad de las redes neuronales artificiales como modelos hidrológicos [1], [2], y [3], los cuáles han demostrado que los modelos basados en redes neuronales obtienen resultados satisfactorios, usando inclusive únicamente datos de caudal y precipitación, los mismos que fueron tratados como series de tiempo.

En ese sentido, el presente artículo hace uso de la versatilidad de las redes neuronales, tomando en cuenta la influencia de los parámetros de configuración, a un enfoque espacial semidistribuido que concluye en la construcción de un modelo precipitación-escurrentía a escala diaria.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

II. METODOLOGÍA

A. Información Básica

El modelo propuesto requiere información climática y de descargas; sin embargo, con el fin de definir las Unidades de Respuesta Hidrológica, base del sistema semi distribuido, se requirió data adicional, tal como la topografía y el tipo de suelo. En ese sentido, la data utilizada se resume a continuación:

Topografía

La topografía se obtiene de la Misión Topográfica Radar Shuttle, a través de la interfaz EarthExplorer del USGS, en formatos DEM, cuya resolución de grilla es de 1” x 1” (equivalente aproximado de 30m x 30m), y la fecha de información es del año 2000.

Hidrometeorología

La Autoridad Nacional del Agua (ANA) de Perú ha puesto al servicio del público su Sistema Nacional de Información de Recursos Hídricos (SNIRH), el mismo que ha sido aprovechado para obtener la información pluviométrica, hidrométrica y de temperatura. La información acerca de cada estación, así como la longitud de sus registros se puede revisar en las tablas I y II:

Tabla I
RELACIÓN DE ESTACIONES HIDROMÉTRICAS

Estación	Barrios	Puente Paltashaco	Teodulo Peña	Obrajillo	Malvados
Lat. (S)	5° 17' 0"	5° 7' 0"	5° 11' 0"	11° 27' 10"	5° 4' 0"
Lon. (O)	79° 42' 0"	79° 54' 0"	79° 54' 0"	76° 37' 19"	79° 44' 0"
Altitud	335	564	178	2706	2527
Registro	1974-87	1974-87	1974-87	2005-17	2001-10
Río principal	Río Bigote Huarmaca	Río Gallegas	Río Corrales	Río Chillón	Río Fortaleza
Área de cuenca (km ²)	420.4	140.7	333.1	366.8	1371.2

Tabla II
RELACIÓN DE ESTACIONES PLUVIOMÉTRICAS

Estación	Coordenadas Geográficas		Altitud	Registro
	Latitud	Longitud		
Altamiza	5° 4' 0"	79° 44' 0"	2527	1974-1987
Barrios	5° 17' 0"	79° 42' 0"	335	1974-1989
Canchaque	5° 22' 0"	79° 36' 0"	1550	1974-1987
Chalaco	5° 2' 12"	79° 47' 30"	2296	1974-1987
Frías	4° 56' 0"	79° 51' 0"	3100	1974-1990
Huancabamba	5° 15' 0"	79° 43' 0"	1954	1974-1990
Morropón	5° 11' 40"	79° 58' 16"	128	1974-1987
Palo Blanco	5° 10' 22"	79° 33' 49"	2736	1974-1988
Paltashaco	5° 7' 0"	79° 52' 0"	1210	1974-1988
Pasapampa	5° 7' 0"	79° 36' 0"	2339	1974-1989

Estación	Coordenadas Geográficas		Altitud	Registro
	Latitud	Longitud		
Santo Domingo	5° 2' 0"	79° 53' 0"	1475	1974-1991
Chamana	10° 12' 26.6"	77° 33' 31.41"	1260	1985-2017
Chiquián	10° 8' 51.47"	77° 9' 34.18"	3386	1997-2017
Cotaparaco	9° 59' 1"	77° 35' 1"	3170	2000-2010
Malvas	9° 55' 38.02"	77° 39' 18.7"	2979	1989-2017
Mayorarca	10° 9' 27.91"	77° 20' 56.01"	3335	1989-2017
Ocos	10° 24' 16.46"	77° 24' .9"	3249	1987-2017
Huaros	11° 24' 0" S	76° 34' 0" O	3585	1989-2017
Milloc	11° 34' 17" S	76° 21' 0" O	4384	1986-2017
Pariacancha	11° 23' 0" S	76° 30' 0" O	3842	1969-2017
Obrajillo	11° 27' 9.50" S	76° 37' 19.40" O	2706	2004-2017
Yantac	11° 20' 0" S	76° 24' 0" O	4617	1982-2017

Tipo de Suelo

A través del Mapa de Suelos del Perú, publicado por el MINAGRI, y elaborado por el ex-INRENA en el año 2000. Obtenido a través de un documento en formato .pdf, con una dimensión de 2362x3245 píxeles y resolución de 96ppp, tanto horizontal como vertical.

B. Área de Estudio

El área de estudio comprende 5 subcuencas de la costa peruana, siendo estas las subcuencas Barrios, Puente Paltashaco y Teodulo Peña pertenecientes a la cuenca del río Piura; la subcuenca Obrajillo, a la cuenca del río Chillón; y por último la subcuenca Malvados, a la cuenca del río Fortaleza. La selección de las áreas de estudio fue de acuerdo con la disponibilidad de datos, específicamente la información hidrométrica, la cual debe ser a nivel diario conforme el alcance del estudio.

Tabla III

PARÁMETROS GEOMORFOLÓGICOS DE LAS SUBCUENCAS DE ESTUDIO

	Unid.	Barrios	Malvados	Obrajillo	Puente Paltashaco	Teodulo Peña
Área	km ²	420.4	1371.2	366.8	140.7	333.1
Perímetro	km	126.7	247.4	141.4	78.6	113.1
Altura media	m	1802.0	3087.8	4424.3	1809.5	1537.5
Longitud de cauce	km	33.2	59.3	44.8	24.0	34.1
Pendiente	%	46.3	49.7	41.3	44.4	47.6
Factor de forma	-	0.4	0.4	0.2	0.2	0.3
Índice de Gravelius	-	1.7	1.9	2.1	1.9	1.7

De estos parámetros podemos observar que la muestra contempla tres (3) cuencas (Barrios, Obrajillo y Teodulo Peña) en un orden de tamaño similar (entre los 300 a 500 km²), una más pequeña (Puente Paltashaco) y otra más extensa (Malvados). Estas cuencas presentan pendientes similares a pesar de poseer alturas medias muy diferenciadas, lo cual indicaría que dicha pendiente sería regular en gran parte del área de las cuencas. Así mismo, el factor de forma indicaría que las subcuencas de Barrios y Malvados no serían cuencas ni alargadas ni ensanchadas, y las cuencas de Obrajillo, Puente Paltashaco y Teodulo Peña son alargadas [4], por lo que estas últimas presentarían un hidrograma con picos menos pronunciados, en comparación con las dos primeras.

Finalmente, resulta necesario representar las cuencas de manera espacial, tal como lo muestra las figuras 1, 2 y 3.

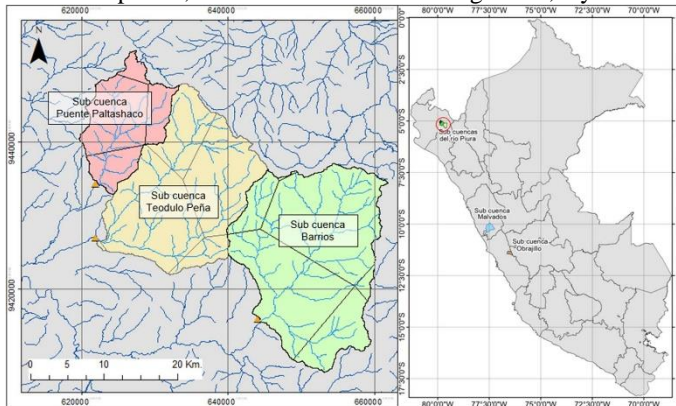


Fig. 1. Ubicación de las subcuencas Barrios, Puente Paltashaco y Teodulo Peña

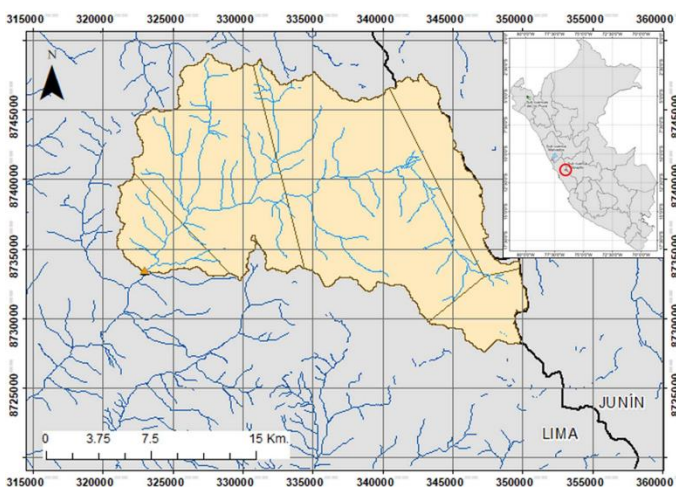


Fig. 2. Ubicación de la subcuenca Obrajillo

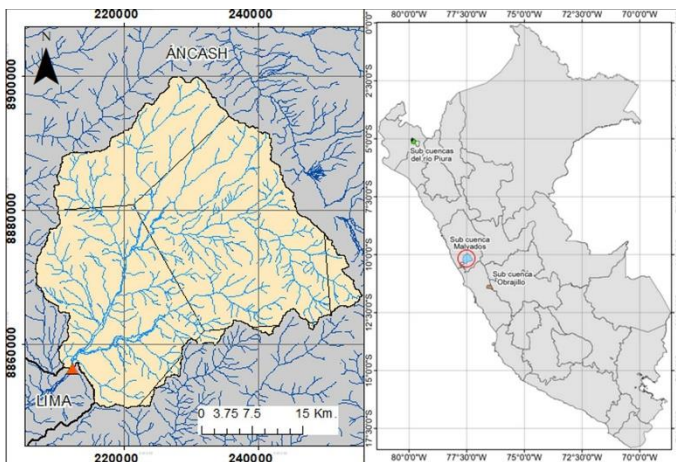


Fig. 3. Ubicación de la subcuenca Malvados

C. Desarrollo de modelo hidrológico

El modelo semidistribuido propuesto se apoya en el concepto de Unidad de Respuesta Hidrológica (URH), definido como las áreas (adyacentes o no adyacentes) con similar comportamiento hidrológico. El comportamiento hidrológico se define en función de un conjunto de variables definidas como

principales y/o características que configuran un conjunto de posibles combinaciones, las cuales son las URH. Para este estudio se tomaron en cuenta dos variables principales: “dirección de pendiente” y “tipo de suelo”, las cuales definen un total de 30 URH, entre todas las 5 subcuencas de estudios, cuyo detalle se encuentra en la Tabla IV.

Tabla IV
RELACIÓN DE UNIDADES DE RESPUESTA HIDROLÓGICA

Estación		Área	Pendiente	Suelo	Dirección de pendiente
Barrios	URH 1	226.11	29.49%	CMe-VRe	N
	URH 2	194.25	66.97%	CMe-VRe	E
Malvados	URH 1	171.45	54.87	LPd-R	N
	URH 2	70.34	54.08	LPq-R	E
	URH 3	68.37	30.94	LPe-CMe	S
	URH 4	170.8	55.48	LPd-R	O
	URH 5	58.03	50	LPq-R	N
	URH 6	53.14	30.23	LPe-CMe	E
	URH 7	240.07	54.96	LPd-R	S
	URH 8	76.28	52.53	LPq-R	O
	URH 9	57.83	31.45	LPe-CMe	N
	URH 10	248.04	53.36	LPd-R	E
	URH 11	82.43	53.48	LPq-R	S
	URH 12	74.39	29.7	LPe-CMe	O
Obrajillo	URH 1	11.28	37.21	LPd-R	N
	URH 2	66.12	41.34	RGd-R	E
	URH 3	39.87	44.94	LPd-R	S
	URH 4	51.95	39.47	RGd-R	O
	URH 5	36.2	43.83	LPd-R	N
	URH 6	58.14	40.55	RGd-R	E
	URH 7	23.27	39.76	LPd-R	S
	URH 8	79.96	41.07	RGd-R	O
Puente Paltashaco	URH 1	20.72	43.58	CMe-VRe	N
	URH 2	28.38	45.24	CMe-VRe	E
	URH 3	49.9	45.23	CMe-VRe	S
	URH 4	41.73	43.36	CMe-VRe	O
Teodulo Peña	URH 1	56.42	48.2	CMe-VRe	N
	URH 2	61.12	48.8	CMe-VRe	E
	URH 3	112.64	47.74	CMe-VRe	S
	URH 4	102.91	46.46	CMe-VRe	O

Definición de parámetros

Por la naturaleza del problema, que es el realizar un modelo de precipitación-escorrentía, la red a desarrollar será del tipo aprendizaje supervisado. Así mismo, tomando en cuenta las

redes aplicadas en las referencias [1], [2] y [3], se hará uso de una red multicapa.

Definido el esquema general, el presente estudio presenta como alcance tres (3) tipos de redes neuronales, la Red Neuronal Prealimentada, esta es del tipo perceptrón multicapa; y la Red NARX (por sus siglas en inglés de “Nonlinear Autoregressive Exogenous”), que se caracteriza por considerar a los datos como series de tiempo, incluyendo así los datos de salida (para el presente estudio es el caudal) de ciclos de tiempo anteriores como nuevos datos de entrada, en sus dos variantes, la red NARX abierta con arquitectura de series en paralelo, donde los datos históricos de caudal forman parte de los datos de entrada, y la red NARX cerrada con arquitectura en paralelo, donde los caudales obtenidos forman parte de los datos de entrada.

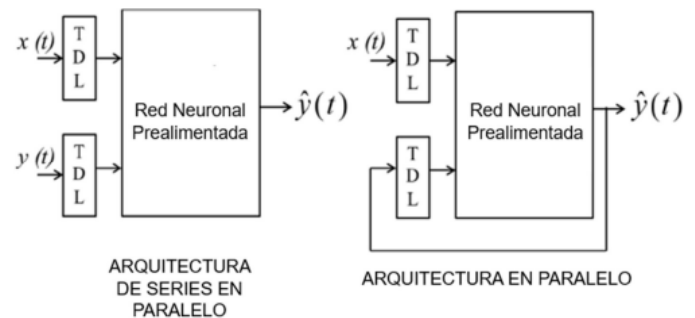


Fig. 4. Esquema de la red neuronal NARX

Los parámetros que se tienen en cuenta para el análisis de la configuración son los datos de entrada, que se permutarán combinaciones entre Precipitación, Temperatura media y Evapotranspiración potencial; el tiempo de retardo, denominado así a la cantidad de ciclos temporales (en este caso la unidad es de días al tratarse de un modelo a escala diaria), y que variará desde un día hasta los cinco días; la arquitectura, referido a la cantidad de capas intermedias, así como la cantidad de neuronas en las mismas; y el periodo de aprendizaje, determinado por la longitud temporal de los datos a usarse durante la etapa de aprendizaje de la red neuronal (para el estudio se aplica un periodo de aprendizaje desde los 2 años hasta los 7 años).

Finalmente, se menciona que el algoritmo de aprendizaje a usarse es el Levenberg Marquardt, la función de desempeño es el Error Medio Cuadrático (MSE) y la función de activación será la tangente sigmoideal, con excepción de la capa de salida, en ese caso será la función lineal.

Preprocesamiento de la información

Consta de tres pasos, primeramente, la Normalización de datos [5], esta se realizará en el intervalo cerrado de -1 a 1, toda vez que es el rango de la función de activación a usar, tal como se observa en las referencias [2] y [3]; en segundo lugar, la generación de series de retardo, toda vez que se hará uso de datos de días anteriores como datos de ingreso, y será según el procedimiento:

Siendo P el registro histórico de la Precipitación normalizada para un periodo de “n” días, y pi la medida de precipitación del día “i” se define:

$$P = [p_1, p_2, \dots, p_n] \quad (1)$$

Entonces se define la matriz P^t como el conjunto de series con retardo “j”, donde este va de 1 a “t”, siendo además la serie de retardo definida como P_j^t.

$$P^t = [P_0^t, P_1^t, \dots, P_t^t] \quad (2)$$

$$P_j^t = [p_{j+1}, p_{j+2}, \dots, p_{n+j-t}] \quad (3)$$

Este procedimiento se aplica para cada dato de ingreso, T para temperatura, ET para evapotranspiración, Q para caudal histórico y Q̂ según corresponde, por lo que la red neuronal con un retardo “t” se podría expresar de la siguiente manera.

$$\hat{Q}_t^t = \text{red}(P^t, T^t, ET^t) \quad (4)$$

Finalmente, se procede a separar las series para las fases de aprendizaje y testeo, siendo la finalidad de este último ser el periodo de evaluación del performance de la red neuronal.

Post procesamiento de la red

Las funciones para medir el desempeño de la red neuronal son las siguientes:

Función de Error Medio Cuadrático (MSE), cuyo valor óptimo es cero, y se define como:

$$MSE = \frac{\sum(\hat{q} - q)^2}{n} \quad (5)$$

Eficiencia de Nash Sutcliffe (NSE), cuyo valor óptimo es uno, y se define como:

$$NSE = 1 - \frac{MSE}{\sigma(q)} \quad (6)$$

Y la Diferencia de Picos (DP), cuyo valor óptimo es cero, y se define como:

$$DP = |\text{máx}(q) - \text{máx}(\hat{q})| \quad (7)$$

III. RESULTADOS

A. Resultados Generales

Tomando en cuenta los resultados de las 5 cuencas de estudios se obtienen lo siguiente:

De acuerdo con el tipo de red, se observa que las redes NARX presentan un mejor desempeño según el NSE, a diferencia de la DP, donde no se observa diferencia sustancial en la tabla V.

	NSE (promedio)	DP (promedio)
Red Prealimentada	0.44	78.53
NARX Abierta	0.67	75.16
NARX Cerrada	0.68	78.33

De acuerdo con el periodo de entrenamiento, se observa en la tabla VI que mientras mayor sea el periodo de aprendizaje, el desempeño se incrementa, sin embargo, a partir de los 5 años de periodo el incremento se hace mucho menor.

	2 años	3 años	4 años	5 años	6 años	7 años
NSE (prom)	0.5206	0.5731	0.5998	0.6118	0.6064	0.6081
DP (prom)	88.39	76.04	74.90	75.09	75.47	74.59

De acuerdo con los datos de ingreso, se observa en la tabla VII que el incluir la temperatura o la evapotranspiración, obtenida a partir de Hargreaves, el desempeño según el NSE o de la DP no se ven afectados.

	NSE (promedio)	DP (promedio)
P	0.59	76.07
P + T	0.58	77.96
P + T + ET	0.59	78.21

B. Análisis por cuencas

Considerando un periodo de aprendizaje de 7 años, se han extraído las redes neuronales que han mantenido un mejor desempeño, y se ha procedido a representar gráficamente el caudal medio mensual, caudal máximo mensual y caudal mínimo mensual, contrastado con el caudal observado (Q_{obs}):

Subcuenca Barrios

Se puede observar en las figuras 5.a, 5.b y 5.c que los caudales pico no se encuentran correctamente representados, en específico al año 1983 correspondiente al FEN del 83. En cambio, los caudales mínimos presentan un mejor ajuste.



Fig. 5.a. Caudal medio mensual – Subcuenca Barrios



Fig. 5.b. Caudal máximo mensual – Subcuenca Barrios



Fig. 5.c. Caudal mínimo mensual – Subcuenca Barrios

Subcuenca Malvados

Se puede observar en las figuras 6.a, 6.b y 6.c que los caudales pico no se encuentran correctamente representados, en específico al año 2009, cabe señalar que el evento registrado es

el mayor del periodo total de análisis. En cambio, los caudales mínimos presentan un mejor ajuste.

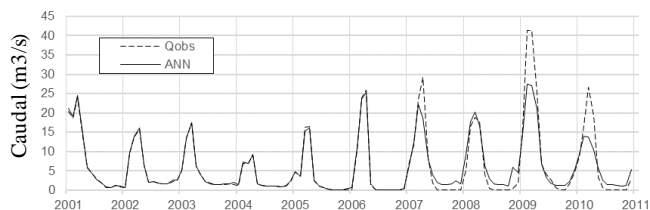


Fig. 6.a. Caudal medio mensual – Subcuenca Malvados

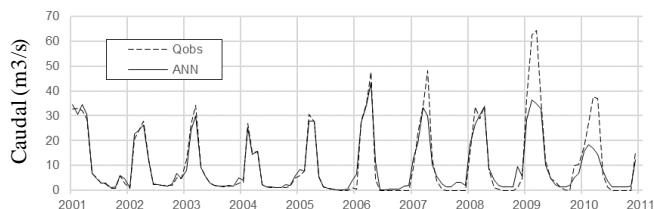


Fig. 6.b. Caudal máximo mensual – Subcuenca Malvados

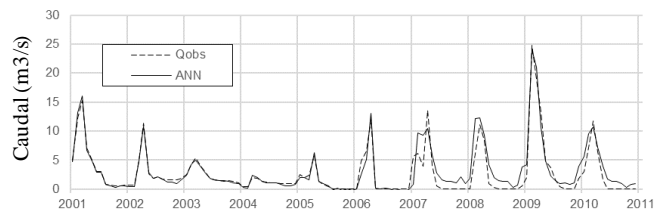


Fig. 6.c. Caudal mínimo mensual – Subcuenca Malvados

Subcuenca Obrajillo

Se puede observar en las figuras 7.a, 7.b y 7.c que los caudales pico, a diferencia de las dos subcuencas anteriores, se encuentran correctamente representados, esto se debe a que no se observa ningún evento extraordinario.

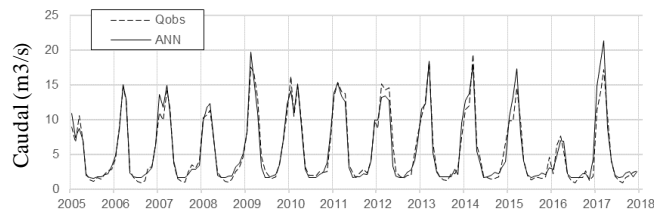


Fig. 7.a. Caudal medio mensual – Subcuenca Obrajillo

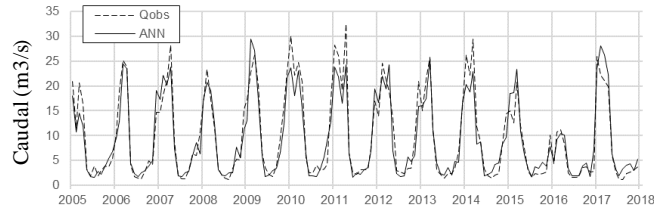


Fig. 7.b. Caudal máximo mensual – Subcuenca Obrajillo

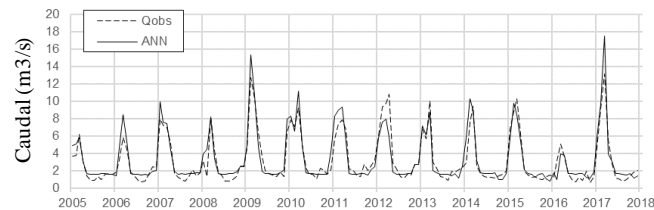


Fig. 7.c. Caudal mínimo mensual – Subcuenca Obrajillo

Puente Paltashaco

Se puede observar en la figura 8.a, el caudal medio mensual se encuentra representado adecuadamente, pero al observar el caudal máximo, figura 8.b, el caudal máximo para el FEN del 83 se encuentra totalmente subvaluado, y que este tiene un orden de magnitud similar al caudal máximo del periodo de entrenamiento (1974 – 1980).

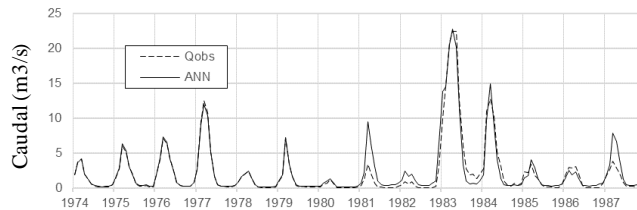


Fig. 8.a. Caudal medio mensual – Subcuenca Puente Paltashaco

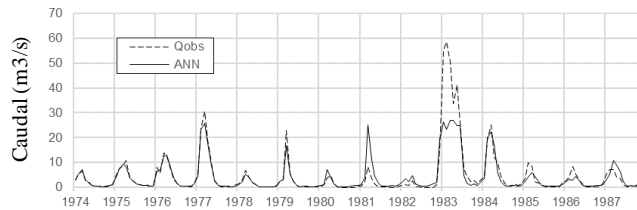


Fig. 8.b. Caudal máximo mensual – Subcuenca Puente Paltashaco



Fig. 8.c. Caudal mínimo mensual – Subcuenca Puente Paltashaco

Subcuenca Teodulo Peña

El comportamiento es similar a la subcuenca Puente Paltashaco, donde el caudal medio mensual (figura 9.a) representa correctamente el FEN del 83, sin embargo, al observar el caudal máximo mensual (figura 9.b) se aprecia que el caudal máximo está subvaluado.



Fig. 9.a. Caudal medio mensual – Subcuenca Teodulo Peña

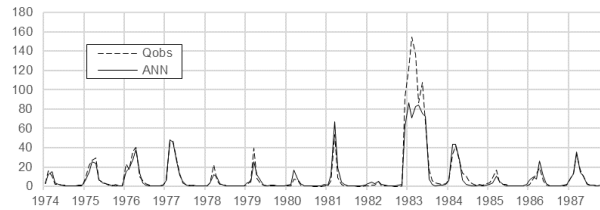


Fig. 9.b. Caudal máximo mensual – Subcuenca Teodulo Peña

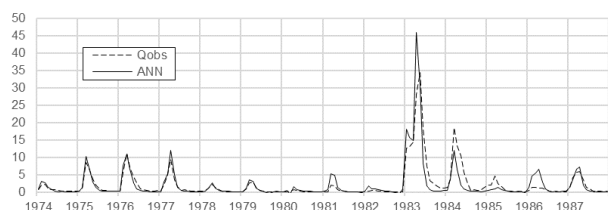


Fig. 9.c. Caudal mínimo mensual – Subcuenca Teodulo Peña

Se ha observado que el caudal máximo, para los eventos del FEN del 83, no se representa adecuadamente como un pico, en cambio se observa un periodo cuasi constante, representado como una línea horizontal (figuras 5.b, 8.b y 9b), esto se debe a que el caudal máximo que el modelo puede llegar a simular se encuentra limitado matemáticamente por la normalización aplicada en esta metodología, así como la función de activación del tipo tangente sigmoïdal.

IV. CONCLUSIONES

Las redes neuronales presentan una capacidad de aprendizaje tal que permite realizar simulaciones con un desempeño que no son menores a 0.5 para la Eficiencia de Nash, siendo un desempeño bueno, inclusive con solo disponer de precipitación y caudal con un periodo de 5 años, siendo 2 años para el aprendizaje y 3 años para la fase de testeo.

Por otro lado, la principal limitante de la red neuronal, de acuerdo a la metodología usada, es la dependencia a sus datos, pero no por su extensión temporal, sino por las diferentes naturalezas que pueda presentarse en dicho periodo de aprendizaje, es decir, que la red neuronal presenta una capacidad limitada al momento de simular un evento extraordinario que no se encuentra en el periodo de aprendizaje.

Para las redes en estudio se observa que para un tiempo de aprendizaje de 3 a 4 años el NSE no se incrementa a más de 0.1 al pasar a los periodos de 7 años, por lo que este periodo es suficiente para que la red pueda aprender de manera óptima el comportamiento de la cuenca, excluyendo los eventos extraordinarios.

REFERENCIAS

- [1] Quiroz K. y Pedrollo O.C. "Artificial neural network for daily rainfall-runoff modelling applied in Urubamba river basin", X Congreso Colombiano de Métodos Numéricos, Cartagena, Colombia, 2015.
- [2] Rajwrkar M.P., Kothiyari U.C. y Chaube U.C. "Artificial neural networks for daily rainfall-runoff modeling". Hydrological Sciences Journal, 47(6), p. 865-877, 2002.
- [3] Singh P. V., Kumar A., Rawat J.S. y Kumar D. "Artificial neural networks based daily rainfall-runoff model for an agricultural hilly watershed". International Journal of Engineering and Management Sciences., 4 (2), p. 108-112, 2013.
- [4] Horton, R. "Drainage Basin Characteristics. Transactions", American Geophysical Union, 13, 350-361, 1932.
- [5] Yu L. & Wang S. y Lai K. "An Integrated Data Preparation Scheme for Neural Network Data Analysis". IEEE Transactions on Knowledge and Data Engineering, 18. 217- 230, 2006.