

## Ajuste de un Modelo Lluvia-Caudal para la Cuenca Media del Río Paute Aplicando el Modelo Neuro-Difuso ANFIS

C. Jordán<sup>1</sup>, D. Medina<sup>2</sup>, A. Zúñiga<sup>3</sup>

Facultad de Ingeniería en Electricidad y Computación

Escuela Superior Politécnica del Litoral

Campus Gustavo Galindo, Km. 30.5 Vía Perimetral

Apartado 09-01-5863, Guayaquil-Ecuador

cjordan@fiec.espol.edu.ec<sup>1</sup>, mrdumax@gmail.com<sup>2</sup>, azuniga@espol.edu.ec<sup>3</sup>

### Resumen

*La construcción de modelos conceptuales en hidrología requiere de considerable tiempo y esfuerzo, pues se necesita contar con una representación muy precisa de las características de la zona de estudio; una alternativa viable son los modelos estocásticos o de caja negra, para los que se requiere de datos abundantes y confiables de ciertas variables como: precipitación, caudal, etc. En este trabajo se presenta el desarrollo y ajuste de dos modelos de caja negra lluvia-caudal para la cuenca media del río Paute, en el área de influencia del embalse Amaluza: ANFIS(M2) y ANFIS(M13); estos modelos se construyeron utilizando un sistema de inferencia fuzzy con base en una red adaptable (ANFIS). Para probar los modelos se realizaron simulaciones para un periodo de 61 meses; los resultados obtenidos se comparan con los valores observados y con aquellos obtenidos mediante la aplicación de un modelo de regresión parcial por mínimos cuadrados (PLS). Para medir la bondad de los modelos se utilizó la medida R<sup>2</sup>. Los resultados demuestran que los modelos ANFIS son adecuados para la representación del proceso lluvia-caudal que ocurre en dicha cuenca, y lo reproducen de mejor forma que el modelo de regresión.*

**Palabras Claves:** Modelo lluvia-caudal, sistemas neurofuzzy, ANFIS, pronóstico de caudales

### Abstract

*The development of conceptual models in hydrology require considerable time and effort because they need a precise representation of its zone of influence characteristics; on the other hand, stochastic or black box models are a viable alternative that require vast amounts of reliable rainfall and runoff data. This paper presents the fitness of two rainfall-runoff black box models for the Middle Basin of the Paute River, at the site of the Amaluza reservoir: ANFIS(M2) y ANFIS(M13). These models were built using adaptable network fuzzy inference systems (ANFIS) and simulations for a period of 61 months were run to test them. Results were compared with both: observed values and those obtained by applying a partial least square regression model (PLS) to the data. Several metrics were used but results are presented with R<sup>2</sup>, because their behavior was basically the same in all cases. Results show that the ANFIS-based models represent appropriately the rainfall-runoff function for the above mentioned basin; it can be said that they were better than the regression model at reproducing such a process.*

**Keywords:** rainfall-runoff model, neurofuzzy systems, ANFIS, discharge forecasting.

## 1. Introducción

El cálculo del caudal de los ríos a partir de datos de precipitación constituye un problema extremadamente complejo; de hecho, el desarrollo de modelos computacionales que representen adecuadamente la respuesta de la cuenca a las precipitaciones es uno de los principales temas de investigación en hidrología. Una clasificación sencilla de los modelos hidrológicos lluvia-caudal consiste de tres tipos: los modelos conceptuales o paramétricos, los modelos métricos, estadísticos o de caja negra, y los modelos distribuidos; estos últimos están fuera del alcance de

este trabajo, y por lo tanto no los mencionaremos de nuevo; en cuanto a los dos primeros, son modelos concentrados que describiremos brevemente a continuación.

Un modelo conceptual es un conjunto de reservorios o depósitos de agua interconectados, que se llenan con la lluvia o la infiltración, y que se vacían por efectos de la evapotranspiración, el drenaje, la escorrentía, etc. El comportamiento de un modelo de este tipo está determinado por un conjunto de ecuaciones que relacionan el contenido de los reservorios (variables de estado) con los flujos en las interconexiones; estas ecuaciones contienen

parámetros que si bien es cierto tienen generalmente una interpretación física directa, a menudo no pueden ser determinados a partir de mediciones de campo debido a que son el resultado de agregar varios procesos. Para efectuar la calibración del modelo, es decir, para calcular los valores óptimos de estos parámetros se requiere disponer de datos suficientes de precipitación y caudal de la cuenca que se modela [16]. Entre los modelos conceptuales más populares se encuentra HYDROLOG [1]; tiene 17 parámetros y otros componentes adicionales para simular el proceso hidrológico; generalmente produce mejores resultados que otros modelos conceptuales que fueron muy populares entre los años 60 y 70, como por ejemplo: el modelo STANFORD (1966) o el modelo SACRAMENTO (1977). MODHYDROLOG es una variante de HYDROLOG que con solo 28 parámetros produce resultados similares a los de su predecesor. También podemos mencionar en este breve resumen al modelo *SIMHYD*, que tiene 7 parámetros y es una versión aún más simplificada de los dos modelos anteriores; *SIMHYD* fue producto de un estudio en el que se aplicó el modelo MODHYDROLOG a 28 reservorios en Australia; a partir de los resultados los investigadores concluyeron que era suficiente usar nueve parámetros, o menos, para obtener estimaciones adecuadas del caudal [2].

El problema con los modelos conceptuales radica, en primer lugar, en que la estructura del modelo es muy subjetiva, pues en gran medida queda determinada por el conocimiento que el modelador tenga del sistema hidrológico [3]; así, dependiendo de quién sea el diseñador se pueden obtener diferentes modelos para una misma cuenca. Además, el número de parámetros del modelo conceptual afecta su precisión al efectuar predicciones; en principio, sería razonable esperar que a mayor número de parámetros la precisión del modelo sea mayor, pero también ocurre mayor incertidumbre. Finalmente, un problema adicional con los modelos conceptuales ocurre cuando existe poca información hidrológica disponible sobre la cuenca que se modela.

Los modelos métricos o estadísticos son modelos de caja negra, en cuyo desarrollo no se necesita ningún conocimiento de la cuenca. Tanto la estructura como los parámetros del modelo se determinan a partir de las series de tiempo de entrada y de salida; así, el modelo se basa enteramente en los datos. La función de transferencia que resulta es característica de la cuenca que se modela y no necesariamente tiene interpretación física alguna. Los parámetros de un modelo de caja negra se calibran utilizando mecanismos computacionales específicos para cada técnica empleada, por ejemplo: los procedimientos de aprendizaje de las redes neuronales artificiales (RNA's). En estos últimos años se han desarrollado modelos conocidos como Sistemas *Neurofuzzy*, que integran los mecanismos de aprendizaje de las RNA's y el poder de interpretación lingüística de los sistemas

difusos; estos sistemas han ganado gran popularidad debido a los excelentes resultados que han mostrado en problemas de pronósticos de series de tiempo e identificación de sistemas.

La literatura sobre aplicaciones de las RNA en el desarrollo de modelos lluvia-caudal es muy extensa, como se muestra en las referencias de [4,5,6]; sin embargo, el número de publicaciones donde se aplica la red ANFIS (*Adaptive Network-based Fuzzy Inference System*) a la simulación de dicho proceso es muy pequeño. En [8] se presenta un estudio sobre la aplicabilidad de la red ANFIS para hacer pronósticos y simulación de caudales en la cuenca del río Sieve (Italia) mediante el uso de una combinación de la red ANFIS [7] y una estructura de entradas similar a un modelo ARX (modelo ARXFIS); el modelo consiguió reproducir el comportamiento del hidrograma del río analizado, y, pudo simular eficientemente los valores pico del caudal. En [9], se analiza la viabilidad del uso de sistemas *neurofuzzy* en la predicción de caudales a partir de datos obtenidos de un simulador experimental de precipitaciones; los resultados obtenidos con la red ANFIS (con muy pocas épocas de entrenamiento) fueron mejores que los que se obtuvieron con modelos de redes neuronales. En [10] se realiza el pronóstico de caudales en el embalse Hirakud (India) para diez días, utilizando como entradas a la red ANFIS una combinación de datos de precipitación y caudal de etapas anteriores, habiéndose logrado buenos resultados si se comparan con los valores de caudal reales. Los resultados obtenidos en los trabajos citados permiten formular la hipótesis de que en el ajuste de modelos lluvia-caudal mediante los sistemas *neurofuzzy* se puede lograr mejores resultados, o por lo menos similares, si los comparamos con los obtenidos mediante otros métodos.

El objetivo de este trabajo es presentar un modelo hidrológico lluvia-caudal de la cuenca media del Río Paute; se trata de un modelo de caja negra construido mediante la aplicación de un sistema de inferencia *fuzzy* con base en redes adaptables (ANFIS). Este modelo servirá para realizar el pronóstico de los caudales afluentes al embalse Amaluza (Central hidroeléctrica Paute-Molino) considerando los posibles escenarios futuros de precipitación bajo los efectos del calentamiento global; este modelo forma parte de un proyecto de investigación que está siendo desarrollado por los autores, y cuyo objetivo es evaluar los efectos del cambio climático en la operación futura del sistema hidroeléctrico Paute, el complejo de generación de energía más importante del país.

El resto de este trabajo está organizado como sigue: en la sección 2 se revisa brevemente la estructura y el proceso de aprendizaje del sistema de inferencia *fuzzy* en la red ANFIS, en la sección 3 se describen las características hidrológicas de la cuenca del río Paute en la zona de emplazamiento del embalse Amaluza, en la sección 4 se explica cómo fueron realizados los ajustes para el modelo lluvia-caudal a utilizar, en la

sección 5 se presentan los resultados obtenidos al modelar dicha cuenca con la red ANFIS, en la sección 6 se presentan las conclusiones. Finalmente, en la sección 7 se presenta una perspectiva de nuestros trabajos futuros

## 2. Redes adaptativas basadas en sistemas de inferencia fuzzy

Los principios de la lógica *Fuzzy*, difusa o borrosa, fueron desarrollados por Lofti A. Zadeh, de la Universidad de California en Berkeley, en la década de los 60's. De acuerdo con Zadeh, "la lógica *fuzzy* es un paradigma de la Inteligencia Computacional que intenta representar formalmente el razonamiento aproximado" [12]. Desde su aparición ha ganado gran popularidad entre los científicos y ha sido exitosamente aplicada en problemas de ingeniería. Los sistemas de inferencia *fuzzy* son sistemas que funcionan con base en reglas; consisten de 3 componentes: un conjunto de valores *fuzzy* que corresponden a los valores lingüísticos que pueden tomar las variables, y que son definidos mediante funciones de membresía o pertenencia, un conjunto de reglas *fuzzy* del tipo *si-entonces*, y un procedimiento que permite combinar dichas reglas y producir las salidas del sistema.

Debido a la capacidad que tienen los sistemas *fuzzy* para ser combinados con otros paradigmas de la inteligencia computacional, ha sido posible desarrollar esquemas que combinan el poder de interpretación lingüística de los sistemas de inferencia *fuzzy* (SIF's) con la capacidad de aprendizaje que poseen las redes neuronales artificiales (RNA's). Estos sistemas, conocidos como sistemas híbridos *neurofuzzy* (SNF's), se caracterizan por representar a los sistemas de inferencia *fuzzy* mediante arquitecturas paralelas, distribuidas e interconectadas que permiten el uso de los paradigmas de aprendizaje de las RNA's.

Bajo estas consideraciones, Roger Jang [7] desarrolló el modelo basado en la red ANFIS (*Adaptive Network-based Fuzzy Inference System*), que utiliza una red adaptable para representar un sistema de inferencia *fuzzy* de tipo Takagi-Sugeno (T-S) de primer orden [13], que es el más eficiente desde el punto de vista computacional. Una red adaptable es una estructura interconectada de capas múltiples, un grafo, donde cada nodo realiza una función específica (función de nodo) sobre sus señales de entrada; de esta manera, la función de nodo puede variar de un de un vértice a otro del grafo [13]; los arcos de la red solo indican la dirección en que se propagan las señales y no tienen pesos asociados. Las redes neuronales artificiales son un caso especial de las redes adaptables. El sistema de inferencia T-S utiliza reglas *fuzzy* de la siguiente forma:

Si  $x \in A_1$  y  $y \in B_1$ , entonces  $f_1 = p_1 x + q_1 y + r_1$

Si  $x \in A_2$  y  $y \in B_2$ , entonces  $f_2 = p_2 x + q_2 y + r_2$

Donde  $x$  y  $y$  son las variables de entrada al sistema de inferencia,  $A_i$  y  $B_i$  corresponden a los valores lingüísticos que pueden tomar dichas variables, respectivamente;  $f_i$  corresponde a la salida de la regla  $i$ -ésima, y  $p_i$ ,  $q_i$  y  $r_i$  son los coeficientes de la función de salida; lineal para el caso especial de inferencia tipo T-S.

La salida de todo el sistema de inferencia se obtiene calculando la media ponderada de las salidas de las reglas, para cada una de las cuales se ha definido un nivel de activación  $\bar{w}_i$  (firing strength) que sirve como peso en el cálculo de dicha media ponderada. En la figura 1 se ilustra gráficamente un sistema de inferencia *fuzzy* del tipo T-S.

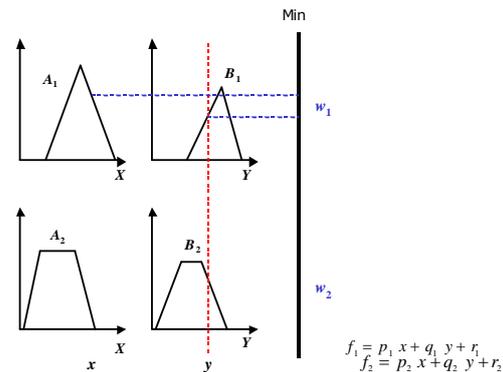


Figura 1. Esquema del Sistema de inferencia T-S [7].

El sistema de inferencia de tipo T-S de la figura 1 es equivalente a una red adaptable de 5 capas que se muestra en la figura 2, donde vemos dos tipos de nodos; los círculos representan nodos sin parámetros, mientras que los cuadrados representan nodos con parámetros, los mismos que se calcularán durante el proceso de aprendizaje; es similar a una red neuronal y por consiguiente se pueden aplicar los paradigmas de aprendizaje de las RNA's.

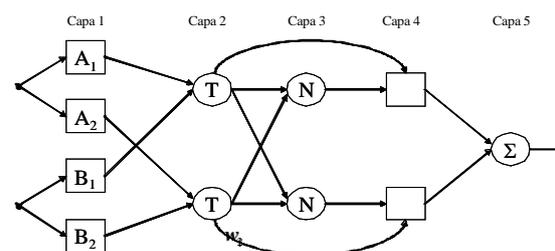


Figura 2. Red adaptable equivalente al sistema de inferencia T-S [7].

Desde la izquierda en la figura 2, los vértices de la primera capa de la red contienen los llamados parámetros del antecedente, que definen las funciones de membresía  $A_i$  asociadas a los valores lingüísticos o *fuzzy* que las variables de entrada pueden tomar; cada salida de esta capa  $O_i^1$  representa el grado al que  $x$  pertenece a  $A_i$ :  $\mu_{A_i}(x)$ .

$$O_i^1 = \mu_{A_i}(x) \quad (1)$$

En la segunda capa, cada nodo se representa con una T y calcula el grado de activación  $w_i$  de cada regla; la T significa que se aplica una t-norma (operación equivalente al operador AND de la lógica binaria) entre las funciones de pertenencia de  $x$  y  $y$ :

$$O_i^2 = w_i = \mu_{A_i}(x) \square \mu_{B_i}(y), i = 1, 2 \quad (2)$$

En la capa 3, cada nodo se representa con una N, y realiza la normalización de los grados de activación de las reglas con respecto a la suma de los grados de activación de todas las reglas; la salida  $O_i^3$  es igual al grado de activación normalizado  $\bar{w}_i$  de cada regla:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1, 2. \quad (3)$$

En la cuarta capa, la salida  $O_i^4$  corresponde al producto entre el grado de activación normalizado de la regla  $i$  y la función de salida  $f_i$  que aparece en el consecuente de dicha regla; donde  $p_i$ ,  $q_i$  y  $r_i$  (parámetros del consecuente del sistema de inferencia) son los parámetros asociados al nodo  $i$ :

$$O_i^4 = \bar{w}_i f_i = p_i x + q_i y + r_i, i = 1, 2. \quad (4)$$

En la última capa se calcula la salida de la toda red  $O^5$  como la suma de todas las señales de entrada a dicho nodo:

$$O^5 = \sum_i \bar{w}_i f_i, i = 1, 2. \quad (5)$$

Así, la red ANFIS quedará definida por cierto número de parámetros asociados a los nodos de las capas 1 y 4, que se conocen como parámetros del antecedente y del consecuente, respectivamente; el número de los parámetros del antecedente dependerá de cuantas variables de entrada tenga el sistema, del número de valores *fuzzy* o lingüísticos que cada una de estas variables pueda tomar, y de la forma que tenga la función de membrecía que se escoja para cada uno de estos valores; por otro lado, el número de los parámetros del consecuente será igual a tres veces el número de reglas que tenga el sistema de inferencia *fuzzy*, puesto que la salida de cada regla está dada en función de los parámetros ( $p_i$ ,  $q_i$  y  $r_i$ ). En la tabla 1 se muestran algunas funciones de membrecía típicas y el número de parámetros que su definición requiere.

**Tabla 1.** Tipos de funciones de pertenencia y cantidad de parámetros a determinarse.

Función de pertenencia	Número de parámetros
Triangular	3
Gausiana	2
G-bell	3
Sigmoide	2

Ahora bien, cualquiera sea la función que una red ANFIS realice, tendrá que aprenderla; el tipo de aprendizaje es con supervisión, es decir, la red aprende su comportamiento a partir de un conjunto de datos, ejemplos o casos de entrenamiento; cada ejemplo es un par ( $e$ ,  $s$ ), donde ' $e$ ' representa la entrada de la red y ' $s$ ' la salida deseada correspondiente.

El aprendizaje de la red ANFIS se realiza mediante un procedimiento que podría ser el algoritmo de "*backpropagation*", comúnmente utilizado en el caso de algunas RNA's, que efectúa una búsqueda de los valores de los parámetros que minimizan una función error; esta función suele definirse en términos de la suma de los cuadrados de las desviaciones de los valores calculados de la salida de la red, con respecto a los valores deseados u observados. Esta búsqueda se realiza mediante la bien conocida estrategia del gradiente descendiente; sin embargo, al aplicar este método se corre el riesgo de quedar atrapado en un mínimo local, y, consecuentemente, de que el proceso de aprendizaje fracase.

Un procedimiento alternativo para el aprendizaje de las redes ANFIS consiste en una estrategia "*hybrid*" o híbrida, que combina el método de estimación de parámetros por mínimos cuadrados y el algoritmo del gradiente descendiente. En la primera fase se inicializa al azar los parámetros del antecedente (funciones de membrecía), y se aplica el método de estimación por mínimos cuadrados para determinar los parámetros del consecuente ( $p_i$ ,  $q_i$  y  $r_i$ ) manteniendo fijos los parámetros del antecedente. En la segunda fase, en cambio, se utiliza el método del gradiente descendiente para ajustar los parámetros del antecedente mientras se mantienen fijos los parámetros del consecuente, que fueron calculados en la primera fase. Este procedimiento iterativo se continúa hasta que se cumpla con alguna condición de terminación apropiada, que, por ejemplo, podría ser cierto valor de la función error, o también cierto número de épocas; una época es una iteración sobre todo el conjunto de los ejemplos de entrenamiento.

### 3. Descripción del área de estudio

La vertiente oriental de los Andes ecuatorianos posee el mayor potencial hidroeléctrico aprovechable en el Ecuador. Una de las subcuencas de mayor importancia energética en la vertiente oriental es la Subcuenca del Paute (cuenca del Río Santiago), que alberga el Proyecto Hidroeléctrico Paute, cuyos 2017 MW de capacidad instalada total (de los cuales se encuentran actualmente en operación 1075 MW) lo convertirán en el proyecto de generación más importante del Ecuador.

La subcuenca del Río Paute está comprendida entre las coordenadas 3°16'37.36"S, 2°17'55.20"S, 79°21'51.19"W y 78°15'36.54"W, tiene un área aproximada de 6,439 Km<sup>2</sup> y ocupa partes de tres Provincias: Azuay, Morona Santiago y Cañar. De

acuerdo a estudios realizados por el Consejo de Gestión de Aguas de la Cuenca del Paute, CG-Paute, la subcuenca en cuestión se puede dividir en tres zonas climáticas: cuenca alta, cuenca media y cuenca baja [12]; el proyecto Paute se ubica en la cuenca media que está comprendida entre los 500 y 2000 msnm. En la cuenca media la temperatura oscila entre los 18°C y 12°C, con una humedad relativa promedio de 89% [14].

En la figura 3 se muestra la ubicación del embalse Amaluz, que suministra el agua para la generación hidroeléctrica de la Central Paute; además, en el mismo mapa se indica las estaciones meteorológicas M410, M222, M217 y M206 que registran, entre otras variables, la precipitación que cae en la zona cercana al embalse; debido a que la estación M217 se ubica en las laderas del embalse Amaluz, se escogió a esta estación para obtener la información histórica de precipitación que se requiere para construir el modelo lluvia-caudal, objeto de este trabajo.



Figura 3. Ubicación del embalse Amaluz.

La cuenca baja del río Paute ha sido clasificada como un área lluviosa de régimen oriental, cuya precipitación anual media varía entre los 2,000 y 4,000 mm; el período lluvioso ocurre cada año de abril a septiembre, mientras que el período seco se presenta de octubre y marzo. Esta situación afecta las operaciones del sistema hidroeléctrico Paute, precisamente en el mes de mayor demanda de energía en el Ecuador (diciembre); los bajos caudales que se presentan en este período impiden que la central genere energía a plena capacidad.

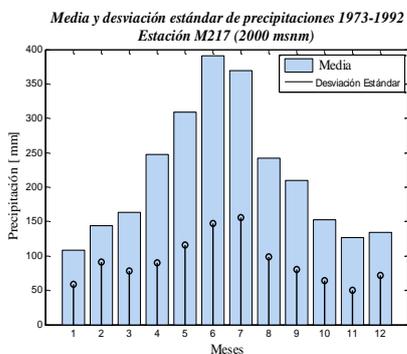


Figura 4. Media multiannual de precipitación en la estación M217 (embalse Amaluz).

En la figura 4 se muestra los valores de la media y la desviación estándar de la serie de caudales medios mensuales para el período comprendido entre los años 1974 y 1992.

El comportamiento estacional de la hidrología se presenta en la figura 5:

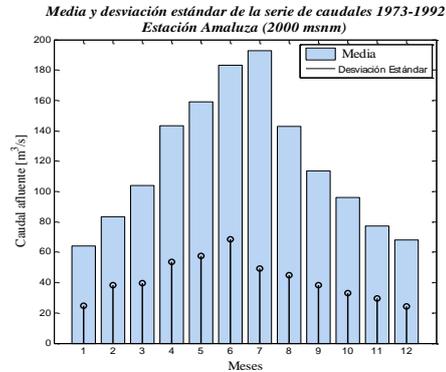


Figura 5. Media multiannual y desviación estándar de caudales afluentes en el embalse Amaluz.

Tabla 2. Estadísticas descriptivas de la serie de precipitaciones y caudales.

	Min	Max	Media	Desviación Estándar	Varianza	Asimetría
Serie de precipitación	19.60	672.20	254.39	136.07	18514.39	0.74
Serie de caudales	26.30	343.00	118.98	61.18	3742.72	0.88

El comportamiento hidrológico de la cuenca armoniza con el de precipitaciones que ocurren en la zona, pues como se observa en las figuras 4 y 5, los meses de mayor caudal corresponden también a los de mayor precipitación. En la tabla 2 se presentan algunas estadísticas descriptivas del caudal mensual en el embalse Amaluz y de la precipitación mensual en la estación meteorológica M217, para el período de 1974 a 1992. Cabe destacar que el caudal medio afluente a dicho embalse (118 m<sup>3</sup>/s) representa aproximadamente el 65% del caudal medio de toda la cuenca del río Paute, de allí la importancia del embalse objeto de nuestro estudio.

#### 4. Ajuste del modelo lluvia-caudal

Los datos que se utilizaron para el ajuste del modelo lluvia-caudal objeto de este trabajo, se seleccionaron de la siguiente forma: los datos de precipitación se tomaron del registro histórico de la estación meteorológica Peñas Coloradas "M217" del INAMHI, situada en la cercanía del embalse Amaluz (ver figura 3); este registro contiene precipitaciones mensuales medias desde 1974 hasta 2004, no obstante, faltan en este registro algunos datos entre 1993 y 1996, y hay inconsistencia en otros desde 1997 hasta el año 2000; por esta razón se escogió el período de mayo de 1974 a noviembre de 1992 para realizar el ajuste del modelo. Los datos que faltaban dentro del período

seleccionado se obtuvieron mediante interpolación utilizando la función *spline* cúbica del programa Matlab V8.0. La información sobre los caudales afluentes al embalse Amaluzá fue proporcionada por la empresa Hidropaute S.A., cuyos registros mensuales abarcan desde el año 1964 hasta el 2008. El total de la información histórica que se utilizó en este proyecto contiene 223 datos de caudales y de precipitación en intervalos mensuales.

En el proceso de ajuste del modelo, los datos anteriores se dividieron en dos subconjuntos: el de entrenamiento y el de pruebas; el conjunto de entrenamiento se utilizó para calcular los parámetros óptimos de la red ANFIS, mientras que el conjunto de pruebas sirvió para evaluar la respuesta de la red frente a datos nuevos; así, 167 datos de precipitación y caudal se utilizaron en el entrenamiento, es decir el 75%; el 25% restante, o sea 56 datos, se usó en la evaluación del modelo.

Determinar cuáles son las variables independientes o predictores es una cuestión esencial que debe considerarse en el proceso de modelar un sistema; en el problema que aquí se aborda, esto no se conoce a priori, aunque se sabe que el caudal de un río en cierto mes depende no solo de la lluvia que cae en dicho periodo, sino también de las precipitaciones y caudales de meses anteriores. En la tabla 3 se listan los 18 modelos que se utilizaron en este trabajo; se especifican también sus salidas y diferentes combinaciones de variables de entrada;  $Q(t)$ , la variable dependiente para todos los modelos, representa el caudal del mes actual,  $P(t)$  es la precipitación en el mes actual,  $P(t-k)$  y  $Q(t-k)$  representan la precipitación y el caudal con retardos de  $k$  meses, para  $k=1, 2, \dots$

Para encontrar la estructura óptima de las variables de entrada del modelo se utilizaron dos estrategias: primero se realizó un análisis estadístico con base en las funciones de correlación parcial entre caudales y precipitaciones, y luego se hizo una búsqueda exhaustiva de la estructura que diera los mejores resultados, de acuerdo con lo propuesto por Jang en [16].

**Tabla 3.** Modelos con diferentes configuraciones de entradas considerados para optimización.

Referencia	Variable Dependiente	Variables Independientes
M0	$Q(t)$	$P(t)$
M1	$Q(t)$	$P(t); Q(t-1)$
M2	$Q(t)$	$P(t); Q(t-1); Q(t-2)$
M3	$Q(t)$	$P(t); Q(t-1); Q(t-2); Q(t-3)$
M4	$Q(t)$	$P(t); Q(t-1); Q(t-2); Q(t-3); Q(t-4)$
M5	$Q(t)$	$P(t); Q(t-1); Q(t-2); Q(t-3); Q(t-4); Q(t-5)$
M6	$Q(t)$	$P(t); P(t-1)$
M7	$Q(t)$	$P(t); P(t-1); Q(t-1)$
M8	$Q(t)$	$P(t); P(t-1); Q(t-1); Q(t-2)$
M9	$Q(t)$	$P(t); P(t-1); Q(t-1); Q(t-2); Q(t-3)$
M10	$Q(t)$	$P(t); P(t-1); Q(t-1); Q(t-2); Q(t-3); Q(t-4)$
M11	$Q(t)$	$P(t); P(t-1); P(t-2)$
M12	$Q(t)$	$P(t); P(t-1); Q(t-1); P(t-2)$
M13	$Q(t)$	$P(t); P(t-1); Q(t-1); P(t-2); Q(t-2)$
M14	$Q(t)$	$P(t); P(t-1); Q(t-1); P(t-2); Q(t-2); Q(t-3)$
M15	$Q(t)$	$P(t); P(t-1); P(t-2); P(t-3)$
M16	$Q(t)$	$P(t); P(t-1); Q(t-1); P(t-2); P(t-3)$
M17	$Q(t)$	$P(t); P(t-1); Q(t-1); P(t-2); Q(t-2); P(t-3)$

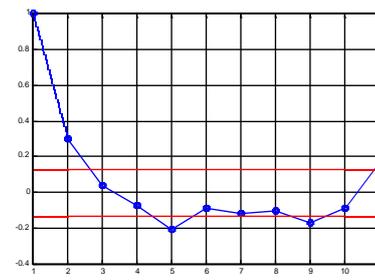
Con el propósito de evaluar el rendimiento de los modelos considerados en este trabajo, se utilizó las 4 métricas siguientes: coeficiente de correlación ( $\rho$ ), coeficiente de determinación ( $R^2$ ), raíz del error medio cuadrático (RMSE) y error medio absoluto relativo (MARE) [15]. Los resultados mostraron que estas métricas son consistentes, es decir, que aquellos modelos con mejor desempeño según con cierta métrica, también lo tienen con relación a las otras tres; por lo tanto, en el resto de este artículo se utilizará únicamente la métrica  $R^2$ , que mide la bondad del ajuste realizado en los parámetros de un modelo; esta métrica ha sido utilizada en otros trabajos similares como se muestra en [3, 4, 6, 8].

#### 4.1. Selección del modelo considerando funciones de autocorrelación.

Con el propósito de determinar de manera racional cuáles deben ser las variables de entrada al modelo, se calculó la función autocorrelación parcial entre  $Q(t)$  y  $Q(t-k)$  para  $k=0, 1, 2, \dots, 11$ , y también la función correlación parcial cruzada entre  $Q(t)$  y  $P(t-k)$  para los mismos valores de  $k$ ; los correlogramas de las figuras 6 y 7 presentan los resultados de estos cálculos, donde las líneas rojas muestran los límites de la banda de confianza del 0.05 aproximadamente. Entonces, si se aplica el criterio que considera que dos variables no están relacionadas estadísticamente, si el factor de correlación parcial entre ellas se encuentra en la banda de confianza del 0.05, concluimos que en este caso las variables independientes o predictores del modelo deberían ser  $P(t), P(t-1), P(t-2), Q(t-1)$  y  $Q(t-2)$ , configuración de entradas que corresponde al modelo M13 de la Tabla 3, al que en adelante se llamará ANFIS(M13).



**Figura 6.** Autocorrelación parcial para la variable caudal, para  $k=1, \dots, 11$ .



**Figura 7.** Correlación parcial cruzada entre  $Q(t)$  y  $P(t-k)$ , para  $k=1, \dots, 11$ .

## 4.2. Método de búsqueda exhaustiva para selección de modelos.

Con el propósito de confirmar los resultados anteriores, se utilizó una estrategia alternativa propuesta por Roger Jang [16]. Este método consiste de dos fases; en la primera se definen varios modelos para la red ANFIS con diferentes configuraciones de variables de entrada, luego se entrena cada uno de estos modelos con unas pocas épocas y se escoge aquel que presente el mejor ajuste. En la segunda fase, se optimiza el modelo seleccionado en la etapa anterior respecto al número de valores *fuzzy* que cada variable puede tomar y al tipo de funciones de membresía que estos valores tienen asociados. De esta manera, con este método se determina cual es la mejor arquitectura para la red ANFIS (tipo y cantidad de funciones de pertenencia por entrada) para el modelo en desarrollo.

El esquema descrito en el párrafo anterior se aplicó a los 18 modelos de la Tabla 3, variando el número de valores *fuzzy* para cada variable de 2 a 5, utilizando diferentes tipos de funciones de membresía: triangular, trapezoidal, gaussiana, etc., pero manteniendo constantes número de épocas: 100, el conjunto de datos de entrenamiento y el algoritmo de aprendizaje: backpropagation. En la Tabla 4 se presentan algunos resultados de esta optimización, donde se observa que el modelo óptimo es el M2, al que en adelante se llamará ANFIS(M2), y cuya configuración de variables de entrada es la siguiente:  $P(t)$ ,  $Q(t-1)$  y  $Q(t-2)$ . Para ANFIS(M2) se obtuvo un valor de  $R^2$  igual a 0.70057, con solo dos funciones de pertenencia de tipo triangular por cada entrada.

**Tabla 4.** Optimización del modelo ANFIS(M2) con respecto al número y tipo de funciones de pertenencia.

Modelo M2		Métricas			
Tipo de Función	# FM	$R^2$ - Eficiencia	Coficiente Correlación	RMSE (m3/s)	MARE (%)
Triangular	2	0.70057	0.83904	30.648	22.235
Trapezoidal	2	0.69376	0.84347	30.994	23.586
Campana Generalizada	2	0.69669	0.83913	30.845	23.500
Curva Gaussiana	2	0.69892	0.83918	30.732	22.891
Combinación Gaussiana	2	0.68879	0.83974	31.244	23.668
PI	2	0.68333	0.83906	31.517	23.617
Diferencia Sigmoidal	2	0.62979	0.83260	34.078	26.268
Producto Sigmoidal	2	0.63997	0.83618	33.606	25.833

## 5. Comparación de resultados

En este trabajo se compara el rendimiento de los modelos lluvia-caudal ANFIS(M2) y ANFIS(M13) arriba descritos, con el de otro modelo que se construyó empleando el modelo de regresión lineal

múltiple denominado Partial Least Square Regression (PLS) [17, 18].

El PLS es un método estadístico que combina características del Análisis de componentes Principales (PCA) con otras del análisis de regresión múltiple. Este método fue introducido por Herman Wold en 1966 [19], y es utilizado ampliamente en la evaluación de daño cerebral.

Para el ajuste de los parámetros del modelo PLS se utilizó el mismo conjunto de entrenamiento que se usó con los modelos ANFIS(M2) y ANFIS(M13), correspondiente a 167 datos de precipitaciones mensuales y caudales medios mensuales. Los predictores o variables independientes del modelo PLS corresponden, en este caso, a las variables que determinan el caudal en la etapa presente, es decir, los valores de  $P(t)$ ,  $P(t-k)$  y  $Q(t-k)$ , para  $k = 1, 2, \dots, 9$ . En la tabla 5 se presenta las variables seleccionadas como predictores en el modelo PLS.

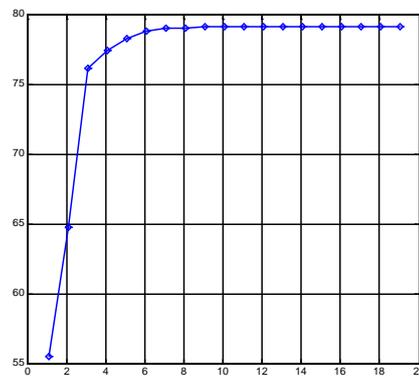
En la figura 8 se muestra como las variables independientes contribuyen a explicar la varianza del caudal; en el eje de las abscisas se representan las variables de entrada al modelo o predictores (tabla 5), y en el eje de las ordenadas se indica el porcentaje de varianza que estas variables pueden explicar.

**Tabla 5.** Variables seleccionadas como componentes del modelo PLS.

N.º de componente PLS	1	2	3	4	5	6	7	8	9	10
Variable	P(t)	P(t-1)	Q(t-1)	P(t-2)	Q(t-2)	P(t-3)	Q(t-3)	P(t-4)	Q(t-4)	P(t-5)

N.º de componente PLS	11	12	13	14	15	16	17	18	19
Variable	Q(t-5)	P(t-6)	Q(t-6)	P(t-7)	Q(t-7)	P(t-8)	Q(t-8)	P(t-9)	Q(t-9)



**Figura 8.** Porcentaje de explicación de la varianza aplicando el modelo PLS.

Con base en esta figura se infiere que el 75% de la varianza calculada para el proceso lluvia-caudal de la cuenca media del Río Paute, a la altura del embalse Amaluzá, se puede explicar con las variables:  $P(t)$ ,  $P(t-1)$ ,  $Q(t-1)$ ,  $P(t-2)$  y  $Q(t-2)$ ; por consiguiente, serán estas variables las que se escojan para ajustar el modelo mediante el método PLS. Nótese que es la misma configuración de variables de entrada para el modelo ANFIS(M13), que se obtuvo mediante un análisis de correlación.

## 5.1. Comparaciones sobre el conjunto de entrenamiento.

Los modelos ANFIS(M2) y ANFIS(M13), que se construyeron utilizando redes ANFIS, fueron entrenados con 167 datos mensuales de lluvia y caudal, correspondientes al 75% del periodo entre 1974 y 1992; en cada caso se hizo una búsqueda de los valores óptimos de los siguientes parámetros: tipo de las funciones de membrecía y número de tales funciones por variable, con el fin de maximizar la medida  $R^2$ ; se utilizó el algoritmo de aprendizaje por backpropagation, y en cada caso el número de épocas fue de 100. Los resultados del entrenamiento para los modelos ANFIS(M2) y ANFIS(M13) se muestran en las figuras 9 y 10, respectivamente.

Cada figura es una gráfica de dispersión que presenta los datos del caudal observado versus el valor calculado según los modelos, y permite reconocer si existe una relación lineal entre ellos, lo que se puede verificar por la cercanía de los puntos a la línea de tendencia. Además, si el aprendizaje del método hubiese sido perfecto, los puntos habrían caído sobre la recta.

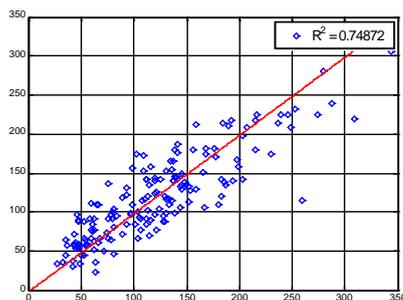


Figura 9. Coeficiente de determinación para el modelo PLS – datos de entrenamiento.

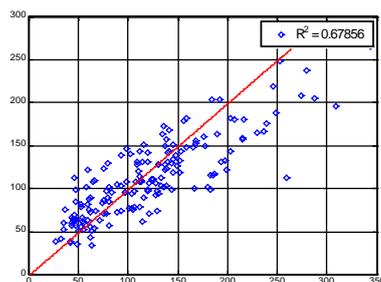


Figura 10. Coeficiente de determinación para el modelo ANFIS (M2) – datos de entrenamiento.

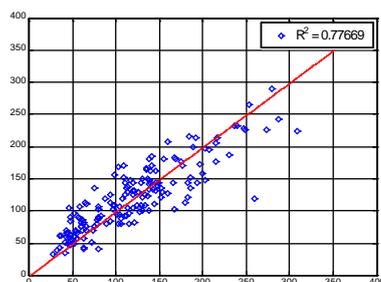


Figura 11. Coeficiente de determinación para el modelo ANFIS(M13) – datos de entrenamiento.

Con los mismos datos se calibró el modelo PLS(M13). La figura 11 presenta los resultados.

De la comparación de los tres modelos: ANFIS(M2), ANFIS(M13) y PLS(M13), en esta fase de calibración o aprendizaje, concluimos que el modelo ANFIS(M13), cuya configuración de variables de entrada se determinó mediante un análisis de correlación, fue el que mejor resultados tuvo: el 77.7% de la variabilidad del caudal  $Q(t)$  se explica por la precipitación en  $t$  y los caudales de los dos meses anteriores; esto significa que la lógica difusa y las reglas *fuzzy* lograron captar el comportamiento de la serie histórica de caudales y la influencia que la precipitación tiene en el proceso hidrológico lluvia-caudal.

## 5.2. Comparaciones en el pronóstico de caudales.

Una vez realizado el aprendizaje de los modelos ANFIS y la calibración del modelo PLS, se realizaron las pruebas de dichos modelos; estas pruebas se realizaron con 56 datos mensuales de lluvia y caudal, correspondientes al 25% restante del periodo entre 1974 y 1992, datos distintos de aquellos que fueron utilizados en la fase de entrenamiento.

La figura 12 recoge dichos resultados; en ella se muestran los hidrogramas para los tres modelos que se desarrollaron en este trabajo y el hidrograma observado o real; como se ve, son muy similares, excepto en algunos pocos picos.

Comparación del pronóstico de caudales basado en la precipitación

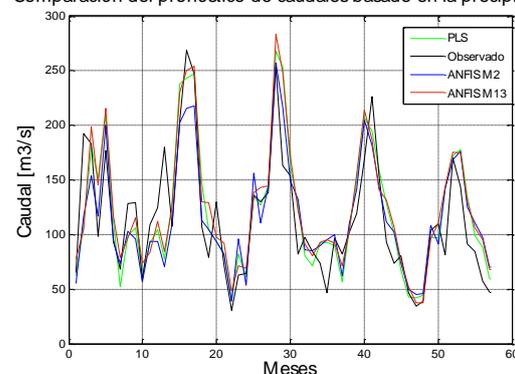


Figura 12. Comparaciones entre los modelos de pronóstico.

En la tabla 6, a continuación, se presentan para los tres modelos en cuestión los valores de las cuatro medidas o métricas que se utilizaron en este estudio.

Tabla 6. Comparación entre los modelos PLS, ANFIS(M2) y ANFIS(M13) frente al pronóstico de caudales.

Modelo	Métricas			
	R <sup>2</sup> - Eficiencia	Coefficiente Correlación	RMSE	MARE
PLS	0.67228	0.84693	32.06274	22.91483
ANFIS(M2)	0.70057	0.83904	30.64805	22.23579
ANFIS (M13)	0.66509	0.85037	32.41299	24.04364

Según estos resultados, el modelo ANFIS(M2) es el que tiene mejor comportamiento en el periodo de pruebas; este fue el modelo cuya configuración de variables de entrada se determinó mediante una búsqueda para optimizar el valor de  $R^2$ , variando el número de variables de entrada, número de valores *fuzzy* que puede tomar cada variable de entrada y tipos de las funciones de membresía para dichos valores.

El valor de  $R^2$  para este modelo fue 0.70057. Para el otro modelo con base en una red ANFIS, el ANFIS(M13), el valor de  $R^2$  fue 0.66509, un 4% menor. Diferencia que podría considerarse como muy pequeña, e interpretar de la siguiente manera: tomando en cuenta que la diferencia entre las configuraciones de las variables de entrada a ambos modelos está en  $P(t-1)$  y  $P(t-2)$ , dadas las características específicas de la cuenca del río Paute a la altura del embalse Amaluza, para calcular el caudal en la etapa presente,  $Q(t)$ , las precipitaciones de los dos meses anteriores o bien no son relevantes, o bien ya están incluidas en los caudales presente y del mes anterior.

## 6. Conclusiones

Este trabajo presentó los resultados del desarrollo de tres modelos de caja negra lluvia-caudal para la cuenca media del río Paute, a la altura del embalse Amaluza: ANFIS(M2), ANFIS(M13) y PLS(M13); los dos primeros modelos se construyeron con base en una red ANFIS, mientras que el último se basó en un modelo de regresión parcial por mínimos cuadrados. La configuración de variables de entrada de ANFIS(M2) se obtuvo por un proceso de optimización, mientras que la de ANFIS(M13) fue el resultado de un análisis de correlación entre las variables lluvia y precipitación.

Según la métrica  $R^2$ , que fue la medida que se utilizó para determinar la bondad del modelo, el ANFIS(M13) fue el mejor de los tres en la fase de entrenamiento, que es cuando el modelo aprende la función que debe realizar; pero, en la etapa de pruebas o evaluación el ANFIS(M2) fue el mejor. Así, en ambos casos, siempre hubo un modelo ANFIS que fue superior al PLS(M13), aunque en general los resultados para los tres modelos fueron similares.

Estos resultados muestran que los modelos construidos en base a una red ANFIS son adecuados para representar la compleja relación no lineal entre caudal y precipitación que existe en una cuenca hidrográfica; además de ser relativamente fáciles de implementar cuando se dispone de los datos necesarios.

En cuanto a los métodos para seleccionar las variables de entrada al modelo, el análisis estadístico permitió obtener un mayor entendimiento de los datos y requirió menos esfuerzo computacional que el otro método, el de optimización, que realizó una búsqueda exhaustiva para encontrar la configuración que produce el mayor valor de  $R^2$ , variando parámetros

como el número de variables, el número de valores *fuzzy* por variable y tipo de las funciones de membresía. Aunque los resultados fueron similares para los modelos que se construyeron de estas dos maneras, es recomendable usar el método que se basa en el análisis de correlación por la razón arriba expuesta.

Finalmente, se ha podido verificar en este trabajo que el desarrollo de los modelos estocásticos o de caja negra requiere mucho menos esfuerzo y tiempo que el de los modelos conceptuales, cuya construcción demanda un conocimiento profundo de las características hidrológicas y meteorológicas de la zona; en el caso de los primeros solo se necesitan datos de lluvia y de caudal que sean confiables y abundantes.

## 7. Trabajos futuros

Este trabajo es parte de un proyecto de mayor alcance, que tiene como objetivo estimar el efecto que tendrá el cambio climático en la operación de centrales hidroeléctricas; en particular, se intentará estimar el impacto que tendrá en la Central Hidroeléctrica Paute, que utiliza las aguas del embalse Amaluza.

Un tema importante de estudio es la degradación del desempeño de los modelos por la consideración de valores pronosticados como base para el pronóstico de largo plazo. Como posibles temas de estudio que se deriven de este trabajo, se pueden mencionar los siguientes:

- Construir un modelo de pronóstico espacial considerando los datos de todas las estaciones meteorológicas que se encuentran en la zona de influencia de una cuenca hidrográfica.
- Estudiar a profundidad el uso de la red ANFIS para la generación de series sintéticas de caudales comúnmente usadas en estudios de planificación de centrales hidroeléctricas.
- Representar el conocimiento que los expertos hidrólogos y meteorólogos tienen en la implementación de modelos lluvia-caudal, para construir un sistema de inferencia *fuzzy* de tipo Mamdani, que parecería ser el más apropiado para representar la experiencia que se tiene del comportamiento de una cuenca hidrográfica, y, consecuentemente, predecir su comportamiento. Una buena base de conocimientos podría compensar el hecho de no haber tomado en cuenta todas las variables independientes relevantes, de las que depende el caudal, y, por otra parte, la falta de informaciones históricas en buen volumen que sean confiables.

## 8. Agradecimientos

Los autores agradecen al Ing. Fausto Alarcón, Coordinador del Proyecto de Adaptación al Cambio Climático del Ministerio del Ambiente del Ecuador, y

al Dr. Santiago Torres, Subgerente de Planificación de Hidropaute S.A., por su aporte con datos e información utilizados en este documento.

Este trabajo se realizó con el financiamiento parcial de la Facultad de Ingeniería en Electricidad y Computación (FIEC), Escuela Superior Politécnica del Litoral (ESPOL), y de la Secretaría Nacional de Ciencia y Tecnología (SENACYT) a través del proyecto PIC-08-00165.

## 9. Referencias

- [1] A. Porter, J. W., "The synthesis of continuous streamflow", Ph.D. Thesis, Monash University, Department of Civil Engineering, Victoria, Australia, 1972.
- [2] Chiew, F. H. S, and McMahon, T.A., "Application of the daily rainfall-runoff model MODHYDROLOG to twenty eight Australian catchments", *Journal of Hydrology* 153, 1994, pp. 383-416.
- [3] Cığizoğ, H.K., Aşkin, P., Öztürk, A., Gürbüz, A., Ayhan, Ö., Yildiz, M., and Uçar, I., "Artificial Neural Networks Models in rainfall-runoff modeling of Turkish rivers", *International River Basin Management Congress Book, March 2007*.
- [4] Kalteh, A.M., "Rainfall-runoff modeling using artificial neural networks (ANNs): modeling and understanding", *Caspian Journal of Environmental Sciences* Vol. 6 No. 1, 2008, pp. 53-58.
- [5] Jeong, D-I and Kim, J-O, "Rainfall-runoff model using artificial neural networks for ensemble streamflow prediction", *Hydrological Processes* 19, 2005, pp. 3819-3835.
- [6] Nagesh, D., and Abhijit, R., "Application of Artificial Neural Networks for Rainfall-Runoff modelling", *Proceedings of 24th National conference on Fluid Mechanics and Fluid Power Vol. 1*, December 1997, pp. D-58.
- [7] Jang, R.J., "ANFIS: Adaptive-network-based fuzzy inference system", *IEEE Transactions on Systems, Man and Cybernetics* 23, 1993, pp. 665-685.
- [8] Gautam, D.K., Holz, K.P., "Rainfall-runoff modeling using adaptive neuro-fuzzy systems", *Journal of Informatics* 3, 2001, pp. 3-10.
- [9] Sy, N. L., "Application of Neuro-Fuzzy Systems to Rainfall-Runoff Modeling Using Field Rainfall Simulator Data", Research Report, University of San Carlos, Cebu City, Philippines, 2002.
- [10] Swain, P.C., Umamahesh, N.V., "Streamflow Forecasting Using Neuro-Fuzzy Inference System", *BALWOIS 2004 Proceedings*, May 2004, pp. 1-10.
- [11] Moreno J., Ceballos, Y., "Pronóstico de generación ideal a escala mensual usando ANFIS", *Revista Avances en Sistemas e Informática* 5, no. 3, 2008, pp. 61-67.
- [12] Zadeh, L. "Fuzzy Logiz, Neural Netowrks, and Soft Computing", *Communications of the ACM* 37, no. 3, 1994, pp. 77-84.
- [13] Takagi, T., Sugeno, M., "Derivation of fuzzy control rules from human operator's control actions", *Proceedings of the IFAC Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis*, July 1983, pp. 55-60.
- [14] Consejo de Gestión de Aguas de la Cuenca del Paute, "Plan Maestro de la Cuenca del Paute VOL I", *Informe del Proyecto de Desarrollo de la Cuenca del Río Paute*, Marzo 2008.
- [15] Firat, M., and Güngör, M., "River flow estimation using feedforward and radial basis neural Networks approaches",
- [16] Jang, J-S. R., "Input Selection for ANFIS Learning", *Proceedings of the IEEE International Conference on Fuzzy Systems*, 1996, New Orleans.
- [17] Abdi, H., "Partial Least Square regression (PLS regression)", in N.J. Salkind ed., *Encyclopedia of measurement and statistics*, Thousand Oaks (CA), pp. 740-744, 2007.
- [18] Alciaturi, C.E., Escobar, M.E., De La Cruz, C., "Partial least squares (PLS) regression and its application to coal analysis", *Rev. Téc. Ing. Univ. Zulia* vol.26, no.3, diciembre 2003, pp.197-204.
- [19] Wold, S, Sjöström, M., Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130.