

DOI: 10.24850/j-tyca-14-01-04

Artículos

**Predicción de caudales mensuales en ríos de cuencas
altoandinas con enfoque de redes neuronales
artificiales. Caso: río Crisnejas, Perú**

**Prediction of monthly flows in rivers of high Andean
basins with an artificial neural network approach. Case:
Crisnejas river, Peru**

Luis Vásquez-Ramírez¹, ORCID: <https://orcid.org/0000-0001-8861-7889>

Luis Félix Vásquez-Paredes², ORCID: <https://orcid.org/0000-0003-1982-4483>

¹Universidad Nacional de Cajamarca, Cajamarca, Perú,
lvasquez@unc.edu.pe

²Consultor, Perú, lfvp505@gmail.com

Autor para correspondencia: Luis Vásquez-Ramírez,
lvasquez@unc.edu.pe



Resumen

Predecir el comportamiento hidrológico en cuencas hidrográficas compuestas por ecosistemas altoandinos que tienen una variedad de climas, con geología compleja, topografía muy variada y suelos con alto contenido de materia orgánica generadoras de una cobertura vegetal muy heterogénea es muy difícil, y si a ello se adiciona la escasez de información hidrométrica en las redes hidrográficas se genera gran incertidumbre al planificar el aprovechamiento del recurso hídrico. La tendencia predominante para la predicción es a través de modelos hidrológicos que relacionan precipitación y escorrentía, los cuales requieren información histórica no disponible en la mayoría de los casos. La aplicación de la técnica de redes neuronales artificiales, en contraste, permite disponer de una metodología adaptable a la información disponible en cada cuenca para analizar la relación entre precipitación y escorrentía, y gracias a su robustez se pueden obtener resultados con gran precisión. El objetivo de esta investigación fue estimar y predecir los caudales promedio mensuales para la cuenca del río Crisnejas, ubicada en la región norte de los Andes peruanos; para ello se contó con registros históricos de 12 estaciones meteorológicas y una estación hidrométrica, utilizando datos de caudal, precipitación, temperatura e índice de vegetación de diferencia normalizada (NDVI), mediante una red neuronal artificial del tipo perceptrón multicapa, con bondad de ajuste del 81 %. Luego, con el registro generado de caudales, se entrenó otra red del tipo recurrente para predecir caudales medios mensuales de ocho años con una bondad de ajuste del 71 %.

Palabras clave: caudales mensuales, redes neuronales artificiales, predicción de caudal mensual.

Abstract

Predicting the hydrological behavior in hydrographic basins composed of high Andean ecosystems that have a variety of climates, with complex geology, highly varied topography, and soils with a high content of organic matter that generate a very heterogeneous vegetation cover, is very difficult, and if it is added the scarcity of hydrometric information in hydrographic networks causes great uncertainty when planning the use of water resources. The predominant trend for prediction is through hydrological models that relate precipitation and runoff, which require historical information that is not available in most cases. The application of the artificial neural networks technique allows a methodology adaptable to the information available in each basin to analyze the relationship between precipitation and runoff. Because of its robustness, results can be obtained with great precision. This research aimed to estimate and predict the average monthly flows for the Crisnejas river basin, located in the northern region of the Peruvian Andes, for which there were historical records of 12 meteorological stations and a hydrometric station, using flow data, precipitation, temperature and normalized difference vegetation index (NDVI), with a multilayer perceptron-type artificial neural network, which achieved a goodness of fit of 81 % in the coefficient of determination. Then with the generated record, another network of the

recurrent type was trained to predict monthly mean flows for eight years with a goodness of fit of 71 %.

Keywords: Monthly flows, artificial neural networks, monthly flow prediction.

Recibido: 22/12/2020

Aceptado: 18/09/2021

Introducción

La estimación de la oferta hídrica es un problema recurrente en la hidrología cuando no se dispone de un adecuado registro de caudales en la cuenca. Para este propósito se dispone de modelos teóricos que se basan en la interrelación de las variables del ciclo del agua y los procesos que ayudan a determinar la cantidad de agua disponible en un punto de interés. La información disponible y requerida en las cuencas determina las características del modelo que se puede aplicar en cada caso; por esto, en ocasiones se hacen simplificaciones o suposiciones respecto a las variables o al ciclo hidrológico, dependiendo de la información

requerida por el modelo. La selección de las variables y la cantidad de datos determina la capacidad predictiva del modelo (Cabrera, 2012).

Torres y Granados (2019) mencionan que tradicionalmente el análisis hidrológico se ha fundamentado en la disponibilidad de información climatológica e hidrológica en una cuenca hidrográfica que permite, junto con el análisis de las condiciones geográficas, geológicas y ambientales, la simulación de fenómenos naturales diversos, como la sequía, inundaciones, crecientes súbitas y disponibilidad de oferta hídrica, entre otros, que resultan ser insumos esenciales para la gestión integral del agua. Sin embargo, cuando se carece de instrumentación en una cuenca no es posible aplicar los protocolos clásicos de modelación, y los hidrólogos se enfrentan al problema de cuantificar el recurso hídrico de formas indirectas y a veces con poco sustento científico. Investigadores como Alipour y Kibler (2019), y Choubin *et al.* (2019) coinciden en que la estimación confiable del caudal, especialmente en cuencas no calibradas, es de suma importancia para la gestión y planificación ambiental, y la predicción del caudal en cuencas no calibradas es necesaria para respaldar las decisiones que se tomen en torno al mejor uso del agua.

Sivapalan y Wagener, citados por Hrachowitz *et al.* (2013), indican que al comienzo del nuevo milenio se había alcanzado una conciencia comunitaria de que las teorías, los modelos y los métodos empíricos hidrológicos eran en gran medida inadecuados para las predicciones en cuencas no calibradas y que además era necesario comprender mejor los vínculos entre la función hidrológica, es decir, la forma en que una cuenca responde a los insumos y la forma, es decir, las propiedades físicas de

una cuenca para abordar adecuadamente el desafío de las cuencas no calibradas.

Así, en las últimas décadas surge la necesidad de encontrar nuevas metodologías capaces de mejorar la precisión en las predicciones de caudales en cuencas no calibradas. Alipour y Kibler (2019) presentan una metodología para la predicción de flujo de corriente bajo el modelo de escasez de datos extrema (SPED), un marco diseñado para la predicción de flujo de corriente dentro de regiones de observación hidrometeorológica dispersa; mientras que Razavi y Coulibaly (2016), y Choubin *et al.* (2019) proponen considerar las características integrales de las cuencas a través de enfoques de modelos múltiples para mejorar la estimación continua del caudal diario en cuencas no calibradas mediante la regionalización, el proceso de transferencia de datos hidrológicos de cuencas hidrográficas calibradas a no calibradas. En la actualidad, muchos investigadores están incluyendo modelos de elevación digital para mejorar la aproximación en los cálculos, es el caso de Althoff, Ribeiro y Neiva-Rodrigues (2021), que presentan una metodología basada en el uso del conjunto de herramientas para el análisis del terreno mediante el modelo de elevación digital (TauDEM), a fin de obtener las variables de entrada para el modelo de regionalización promediadas para el área de captación de cada píxel en la cuadrícula de la red de caudal.

Hrachowitz *et al.* (2013), luego de concluir su investigación titulada “Una década de predicciones en cuencas no calibradas (PUB): una revisión”, encontraron que los principales factores que contribuyen a la incertidumbre predictiva resultante, que fueron identificados por la

iniciativa PUB, incluyen lo siguiente: a) una comprensión incompleta del conjunto de procesos que subyacen a la respuesta del sistema hidrológico y las retroalimentaciones a escala de captación entre estos procesos, lo que frecuentemente resulta en modelos inherentemente irreales con alta incertidumbre predictiva; b) una comprensión incompleta de la heterogeneidad espacio-temporal de múltiples escalas de los procesos en diferentes paisajes y climas, pues la gran mayoría de las pequeñas cuencas de captación en todo el mundo estaban, y siguen estando, sin medir con poca o ninguna información disponible; y c) técnicas de regionalización inadecuadas para transferir la comprensión de los patrones de respuesta hidrológica de ambientes medidos a no medidos debido a la falta de estudios comparativos entre cuencas hidrográficas y la falta de comprensión de los principios físicos que gobiernan una regionalización sólida.

En cuencas pequeñas, o casos en los que se dispone de poca cantidad de datos, o eventos específicos de precipitación, la relación directa entre lluvia y escorrentía se puede determinar mediante métodos de regresión (Osborn & Lane, 1969), deduciendo ecuaciones que puedan relacionar el caudal con la lluvia y/o más variables (USACE, 1971). Estas técnicas dan mayor flexibilidad en cuanto a la información requerida, aunque con mayor número de suposiciones y sin una interrelación conocida entre las variables intervinientes en el proceso, en comparación con los modelos hidrológicos. Además, por la naturaleza del método, la extrapolación de valores es limitada, no se pueden resolver las relaciones

no lineales sin transformar las entradas y es sensible a los valores atípicos.

En contraste con las técnicas y modelos de estimación de caudales descritos, las redes neuronales artificiales (en adelante RNA) presentan ventajas en cuanto a que no es necesario conocer la relación física entre las variables que intervienen en el problema; son robustas (no presentan alta sensibilidad a errores en los patrones de entrada); las variables de entrada pueden adaptarse a los datos disponibles (Delgado, 1998), y dependiendo del tipo de RNA pueden aplicarse en procesos recurrentes para realizar pronósticos de series de tiempo. Para Herrera, Leiva y Romero (2020), en hidrología existen muchos casos donde las redes neuronales se han utilizado para predecir el comportamiento de una variable con base en datos históricos previos y en un conjunto de variables predictoras, pues en su investigación abordaron el problema particular de la reconstrucción de información faltante de las estaciones meteorológicas utilizando RNA.

En las últimas décadas, el uso de redes neuronales en la modelización hidrológica ha aumentado debido a su propiedad fundamental como aproximador universal y parsimónico de funciones no lineales. En el campo de la previsión de inundaciones, los perceptrones de alimentación directa (*feed-forward*) y de tipo multicapa recurrentes (*recurrent multilayer*) han confirmado su eficiencia (Darras, Johannet, Vayssade, Kong-A-Siou, & Pistre, 2018). La gestión sostenible de los recursos hídricos requiere previsión de caudales en tiempos cortos, desafíos hidrológicos que Steyn (2018), y Lama y Sánchez (2020)

proponen enfrentarlos con la aplicación de técnicas de aprendizaje automático tanto para tratar la discontinuidad de los datos como para trabajar con caudales que siguen comportamientos no lineales ni estacionarios. Por su lado, Brenes-Jiménez (2020) puntualiza más la predicción del caudal promedio horario utilizando modelos de *Machine Learning* basados en árboles de decisión, comparando su capacidad predictiva en la estación hidrológica Palmar, ubicada sobre el río Grande de Térraba en la región Pacífico-Sur de Costa Rica.

Para Heras y Matovelle (2021), los métodos computacionales basados en el aprendizaje automático han tenido un amplio desarrollo y aplicación en hidrología, especialmente para sistemas de modelado que no tienen suficientes datos. Dentro de este problema hay series de datos que faltan y que no necesariamente deben descartarse; esto se logra mediante la completación de los mismos, entendiendo que para ello se requiere combinar enfoques o metodologías; en ese sentido se han desarrollado algunas investigaciones que han tenido resultados satisfactorios, como el de Canchala, Alfonso-Morales, Carvajal-Escobar, Cerón y Caicedo-Bravo (2020), que evaluó el desempeño de la combinación de tres enfoques de redes neuronales artificiales (ANN) en el pronóstico de las anomalías de lluvia mensual para el suroeste de Colombia; o el de Farfán, Palacios, Ulloa y Avilés (2020), que propone una técnica híbrida, utilizando las series de tiempo generadas por los modelos individuales como entradas de una nueva RNA. Este enfoque tiene como objetivo aumentar la precisión del flujo simulado mediante la

combinación y explotación de la información proporcionada por modelos físicos y basados en datos.

En el río Crisnejas, ubicado al norte del Perú, se tiene un registro mensual de caudales de 13 años en dos periodos separados por un vacío de datos durante 37 años; sin embargo, sí se cuenta con registros completos de precipitación y temperatura en muchas estaciones climáticas dentro y alrededor de la cuenca. Esta situación es común en cuencas de la costa y sierra peruanas de gran interés en la implementación de proyectos hidráulicos para los que se necesita conocer la disponibilidad de agua. El corto periodo de registro impide una estimación probabilística adecuada de la persistencia de caudales y por esta razón el registro debe ser completado a partir de relaciones lluvia-escorrentía (ANA, 2015). En ese sentido, en el país se aplican frecuentemente modelos hidrológicos para caudales mensuales, tales como el de Témez, de valoración global y para cuencas inferiores a 10 000 km² (Témez, 1977), o el modelo desarrollado por Lutz Scholz en el marco de la Cooperación Técnica de la República de Alemania para el Plan Meriss II, y que es aplicable solo a cuencas de la sierra peruana (Scholz, 1980).

Los modelos hidrológicos antes mencionados requieren de simplificar los datos de precipitación proveniente de las estaciones en un registro promedio dentro de la cuenca, eliminando la variabilidad. Algo similar ocurre con la temperatura y en el caso de Témez se requiere además de estimar la evapotranspiración potencial (ETP) promedio en la cuenca, pero no siempre se cuenta con datos suficientes y se debe optar

por modelos de estimación de ETP a partir de temperaturas. En el proceso de calibración y validación de estos modelos es posible llegar a encontrar valores absurdos en parámetros como la descarga del acuífero, o los coeficientes de retardo y escorrentía, dado que no siempre se puede aplicar en la cuenca de interés o simplemente se tienen deficiencias en los datos de entrada.

Frente a la propuesta de estimación de caudales mensuales mediante los modelos hidrológicos mencionados, las RNA no eliminan la variabilidad de los datos de precipitación provenientes de las diferentes estaciones climáticas, sino que establecen de manera implícita o interna su influencia en los datos de salida. Algo similar ocurre con los datos de temperatura o las variables adicionales que se pueden considerar en el análisis. Además, la calibración no es necesaria, pues la RNA buscará “aprender” por sí sola cómo debe relacionar las entradas para llegar a la salida con el menor error posible (Delgado, 1998). Esto brinda mucha flexibilidad en cuanto a la información de entrada y la calidad de los resultados.

Por lo anterior, el objetivo de esta investigación es aplicar las redes neuronales artificiales (RNA) para estimar los datos de caudal faltantes en el río Crisnejas a partir de datos de precipitación, temperatura y cobertura vegetal cuantificada mediante el índice de vegetación de diferencia normalizada (NDVI) de un año promedio.

Antecedentes

Las redes neuronales artificiales son una técnica computacional inspirada en el trabajo del modelo de neurona biológica y lógica umbral de Warren McCulloch y Walter Pitts en el año 1943; el principio del perceptrón se estableció en 1958 con su limitación a la resolución solamente de problemas separables linealmente, pero no es sino hasta 1975 cuando se propone el algoritmo de propagación inversa o *backpropagation*, y se resuelve tal limitación (Delgado, 1998).

Las investigaciones que aplican directamente las redes neuronales artificiales (RNA) para la solución de problemas hidrológicos complejos han sido cada vez más frecuentes dada la gran cantidad de herramientas computacionales desarrolladas en los últimos años para el entrenamiento de RNA, y sus diferentes algoritmos y tipos. Se presenta a continuación un resumen de los trabajos previos que anteceden a esta investigación.

El *Journal of Hydrologic Engineering* (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000) presenta un artículo en el que se recopilan las posibles aplicaciones de las RNA en diversas ramas de la hidrología, como: lluvia-escorrentía, caudales, aguas subterráneas, calidad de agua y precipitación. Indica que, con un adecuado entrenamiento, las RNA pueden generar resultados satisfactorios para la predicción de los problemas en la hidrología.

Dawson y Wilby (2001) proponen un protocolo para la implementación de redes neuronales artificiales en procesos precipitación-escorrentía y predicción de inundaciones en el que se menciona un proceso de normalización o tipificación de los datos en el rango que es aceptado por la función de activación.

Kalteh (2008) realiza una modelación precipitación-escorrentía y RNA mediante datos de precipitación, temperatura, caudal y tiempo. Su investigación concluyó en una precisión razonable en la estimación del caudal mediante las RNA; además, las señala como herramientas prometedoras no sólo en cuanto a precisión del modelo sino también en la relación aprendida, dado que utilizó métodos de interpretación neuronal para interpretar la conexión entre los pesos de la red.

En su investigación, Laqui (2010) utiliza los datos de precipitación, evapotranspiración y caudal del río Huancané (Perú) para el entrenamiento de una RNA del tipo perceptrón multicapa con el algoritmo de *backpropagation* y compara sus resultados con un modelo de serie temporal estocástica, obteniendo un mejor ajuste con la RNA.

Herrera-Quispe, Yari, Luque y Tupac (2013) utilizaron también RNA del tipo perceptrón multicapa con el algoritmo de Levenberg-Marquardt para generar caudales mensuales estocásticos en la cuenca del río Chili (Perú), en combinación con el modelo estocástico de Thomas-Fiering.

Gomes-Villa-Trinidad (2016) presenta en su trabajo de fin de máster un modelo de predicción de aportaciones mensuales utilizando RNA en la cuenca del río Amambaí (Brasil). Sus conclusiones mostraron

que utilizando 10 neuronas ocultas se pudo obtener mejores resultados que con redes de 15 a 25 neuronas. Además concluye en que las RNA son una alternativa muy eficiente para realizar predicciones de caudales en contraste con el modelo Naive de predicción trivial. Esta investigación también recopila la metodología propuesta por Dawson y Wilby (2001) en forma de protocolo para implementar modelos precipitación-escorrentía con RNA. La investigación también empleó RNA del tipo perceptrón multicapa con el algoritmo de Levenberg-Marquardt.

Materiales y métodos

Propuesta metodológica

Para determinar el caudal mensual histórico en el periodo 1965-2017 y hacer su predicción en un periodo de ocho años se propone el entrenamiento de dos redes neuronales artificiales del tipo perceptrón multicapa y recurrente NAR.

Para la primera red, los patrones de entrenamiento tienen como entrada los siguientes datos:

- Precipitación. Registrada mensualmente (1965-2017) en 12 estaciones meteorológicas ubicadas en la zona de estudio (límite de la cuenca y alrededores).
- Temperatura. Registrada (1965-2017) por 5 de las 12 estaciones anteriores.
- Cobertura del suelo. Cuantificada a partir del NDVI y obtenida a partir de imágenes multiespectrales de cada mes de un año hidrológico promedio.
- Caudales. Registro mensual corto (1965-1976, 2014-2019) utilizado en tres partes: una para el entrenamiento de la red tipo perceptrón multicapa (1968-1976, 2016); otro para la validación de dicha red (2014, 2015, 2017), y otro periodo más corto (2018-2019) para la validación de la predicción hecha con la red tipo recurrente NAR.

El diagrama propuesto en la Figura 1 muestra el proceso seguido para el entrenamiento y la predicción con ambas redes.

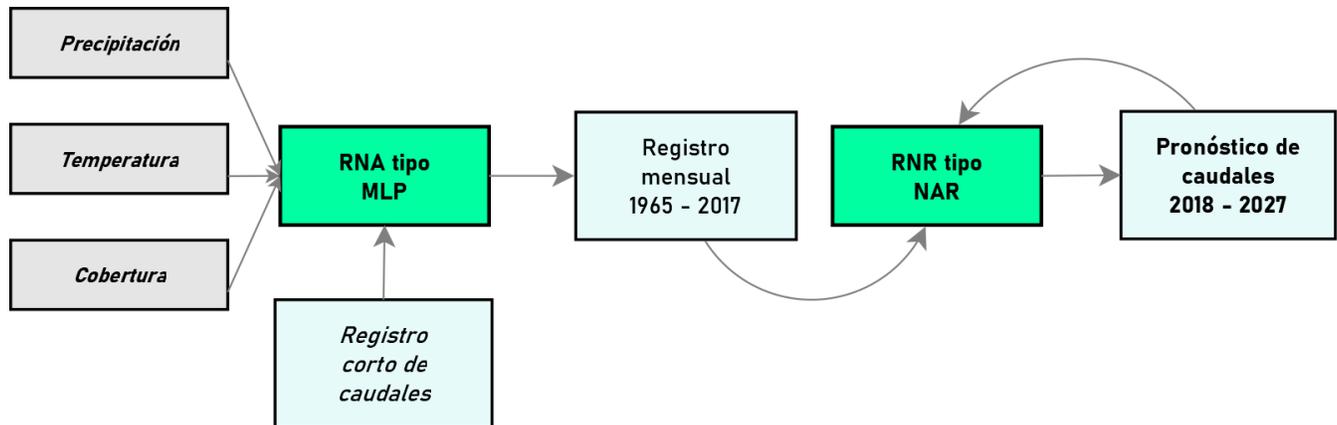


Figura 1. Metodología para la estimación de caudales.

Para la segunda red (RNR NAR) solamente se requieren los datos del caudal estimado con la RNA MLP.

En la actualidad existen muchas herramientas para llevar a cabo el entrenamiento de redes neuronales artificiales de todo tipo, desde lenguajes de programación como Python o R hasta programas con interfaz gráfica como MATLAB. Para este caso, el entrenamiento de la RNA tipo MLP se ha hecho mediante la programación del algoritmo de *backpropagation* en lenguaje VB.net. Para el caso de la RNR tipo NAR se ha utilizado la aplicación *Time Series App* de redes neuronales artificiales de MATLAB 2015.

Balance hídrico

La metodología propuesta para la estimación del caudal mensual (m^3/s) en la cuenca del río Crisnejas parte del planteamiento de las variables más influyentes en el balance hídrico de la cuenca. Según Fattorelli y Fernández (2007), el modelo hidrológico de una cuenca está fundamentado en los procesos que integran las fases del ciclo hidrológico. En una cuenca podemos encontrar diversas variables que se clasifican en entradas (precipitación); salidas (escorrentía, flujo subterráneo, evapotranspiración), y variación del almacenamiento. Todas estas variables se interrelacionan como se muestra en la Ecuación (1):

$$\Delta S = P - Q - G - ET \quad (1)$$

Donde:

ΔS = almacenamiento en mm/año por área de la cuenca.

P = precipitación en mm/año por área de la cuenca.

Q = caudal en mm/año por área de la cuenca.

G = flujo neto de aguas subterráneas hacia el exterior de la cuenca en mm/año por área de la cuenca.

ET = evapotranspiración en mm/año por área de la cuenca.

Al analizar cada una de las variables se observa que el conocimiento de la precipitación es primordial para efectos de estimar el caudal; en este caso se considera independiente de otros factores, y es un dato medido y a considerar en el patrón de entrada de la RNA.

El flujo subterráneo depende de la cobertura, el tipo de suelo y la geología; estos dos últimos se consideran constantes en la escala de tiempo mensual y el periodo global analizado (53 años), por lo tanto, el parámetro a cuantificar viene a ser la cobertura. En tal sentido, se ha propuesto la cuantificación de dicho parámetro a través del NDVI o Índice Diferencial de Vegetación Normalizado según Huete y Tucker (1991), en un año promedio.

La evapotranspiración, según Allen, Pereira, Raes y Smith (2006), es la combinación de dos procesos separados por los que el agua se pierde a través de la superficie del suelo por evaporación y transpiración de la vegetación. Existen muchas ecuaciones o métodos para su estimación. En esta investigación se ha tomado su conceptualización más sencilla. Thornthwaite (1948) plantea la Ecuación (2), que brinda una estimación de la ETP en mm/día:

$$ETP = 16(10 \times T/I)^a \quad (2)$$

Donde:

T = temperatura en °C.

I = índice de calor anual, es función de la temperatura mensual.

a = parámetro en función de I .

De esta manera, la evapotranspiración potencial no necesita ingresar directamente en el modelo, ya que puede ser expresada como función de la temperatura. Su comportamiento también será mejorado a partir del NDVI, dado que, realmente, también es dependiente de la cobertura de la cuenca.

El almacenamiento está relacionado con procesos complejos en los cuales se debe tener en cuenta la cobertura, el tipo de suelo, la geología, la infraestructura y el relieve. Su variabilidad no es significativa en el periodo de tiempo y la escala temporal de la investigación, por lo tanto es una constante.

Finalmente, se propone el modelo conceptual para estimar el caudal mensual con base en la precipitación (P), temperatura (T) y NDVI:

$$Q = f(P, T, NDVI) \quad (3)$$

Red del tipo perceptrón multicapa y algoritmo de retropropagación

Según Isasi-Vinuela y Galván-León (2004), a diferencia del perceptrón simple, el perceptrón multicapa permite resolver problemas no separables linealmente. Este tipo de red está compuesta por varias capas ocultas que permiten tener regiones de decisión. El perceptrón multicapa o MLP (*multi-layer perceptron* se suele entrenar por medio del algoritmo de propagación inversa o retropropagación (*back propagation*), razón por la cual también es conocido con el nombre de red de retropropagación.

Las RNA del tipo perceptrón multicapa (Figura 2) están compuestas de una capa de entrada, una o varias capas intermedias u ocultas, y una capa de salida. Cada una de las neuronas en las capas anteriores se conectan con todas las neuronas de las capas siguientes. La información se propaga en un solo sentido: una vez que la información se presenta en la RNA en la capa de entrada, ésta llega hasta la capa de salida pasando por las ocultas; dicho proceso se denomina *feedforward*.

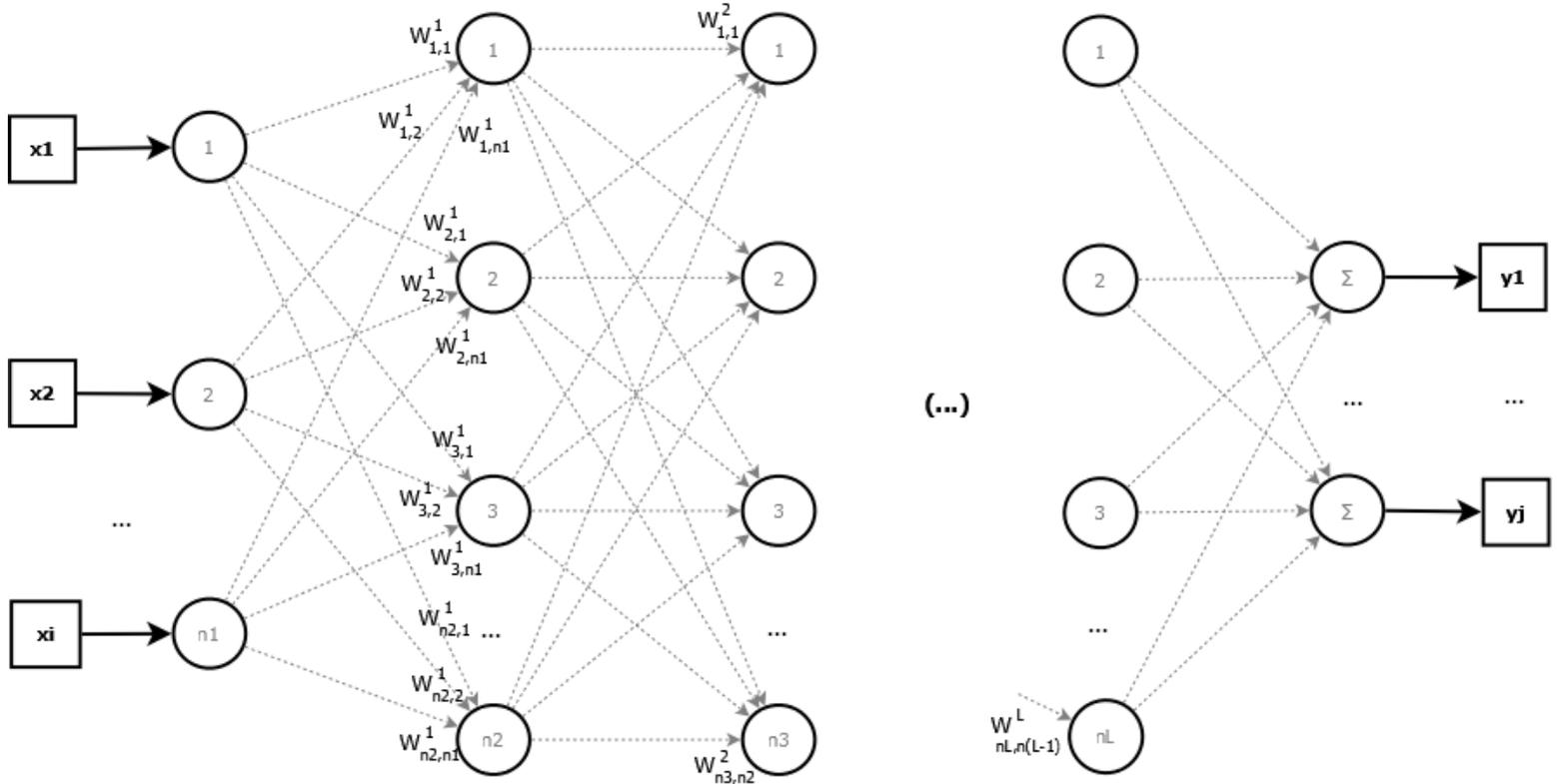


Figura 2. Perceptrón multicapa.

Cada neurona recibe una combinación lineal (sumatoria) de la información afectada por los denominados "pesos" y luego es evaluada por la "función de activación", la misma que genera la entrada para la capa siguiente, tal como se interpreta de la Ecuación (4), según Delgado (1998). Los pesos se ajustan mediante el proceso de entrenamiento, para el cual existen algoritmos como el de retropropagación, que se combina con técnicas de minimización de errores, como el descenso de gradiente, Levenberg-Marquardt, Newton o gradiente conjugado:

$$y = \sigma(\sum W * x + W0) \quad (4)$$

Donde:

y = salida de la neurona.

σ = representa la función de activación (FA); puede ser del tipo tangente, logística, identidad, ReLU, Gaussiana u otras.

x = son las entradas.

W = pesos.

$W0$ = umbral de activación.

Algoritmo de retropropagación

El algoritmo de retropropagación para entrenar una arquitectura MLP (*multi-layer perceptron* o perceptrón multicapa) consta de cinco pasos elementales según Larranaga, Inza y Moujahid (1997), los cuales son:

Paso 1. Establecer aleatoriamente los pesos y umbrales iniciales ($t:=0$, época inicial).

Paso 2. Para cada patrón del conjunto de entrenamiento:

2.1 Ejecutar una fase para obtener la respuesta de la red en el patrón.

2.2 Calcular el error total en la capa de salida.

2.3 Calcular el incremento parcial de los pesos y umbrales debidos a cada patrón de entrenamiento.

Paso 3. Calcular el incremento total actual, extendido a todos los patrones. Se realiza el mismo procedimiento con los umbrales.

Paso 4. Se actualizan los pesos y umbrales.

Paso 5. Se determina el error total y si no es aceptable se vuelven a presentar a la red todos los patrones y se repite el algoritmo desde el Paso 2 hasta obtener resultados satisfactorios ($t:=t+1$, época siguiente).

Blanco (2016) indica que el algoritmo de retropropagación se suele combinar con algún algoritmo de aprendizaje, como la regla delta o el método del descenso de gradiente. Con este último se ha llevado a cabo el entrenamiento de la RNA utilizada en esta investigación.

El esquema propuesto para el entrenamiento se muestra en la Figura 3. Las principales características de la RNA utilizada para la estimación del registro histórico de caudales mensuales del periodo 1965-2017 es:

- Tipo RNA : perceptrón multicapa
- Algoritmo de entrenamiento : propagación inversa
- Algoritmo combinado : descenso de gradiente
- Función de activación única : tangente hiperbólica
- Estructura de la RNA : 7 – 5 – 4 - 1 neuronas por capa

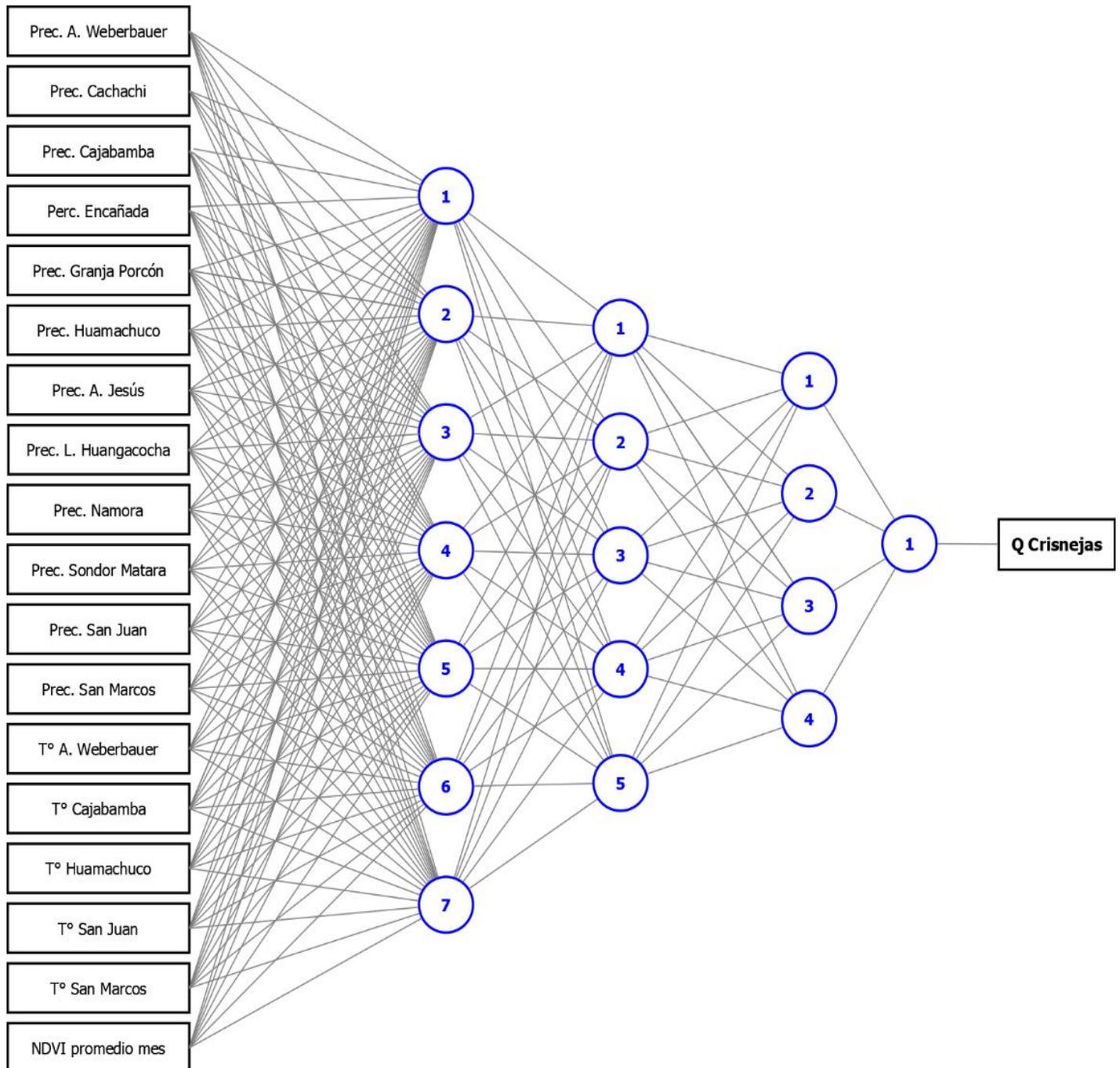


Figura 3. Red neuronal artificial perceptrón multicapa entrenada.

Este esquema se ha obtenido a partir múltiples procesos de prueba y error, siendo el que obtuvo mejores resultados en entrenamiento y en la extensión de los valores no entrenados.

Red del tipo recurrente

Pérez-Ortiz (2002) explica en su tesis doctoral que la forma en que se interconectan las neuronas de una RNA define un grafo dirigido. Si el grafo es acíclico estamos ante el caso más común de una RNA de propagación hacia adelante o *feedforward*, tipo de red en el que se encuentran las RNA tipo perceptrón multicapa vistas anteriormente. En el caso de que la red posea ciclos se denomina red neuronal recurrente. En este tipo de redes los ciclos existentes tienen un profundo impacto en la capacidad de aprendizaje de la red y las hacen más eficientes para el procesamiento de series temporales.

Una red neuronal recurrente (RNR) puede ser de varios tipos. En esta investigación se ha empleado una RNR tipo NAR (*Non linear autoregressive*), también conocida como modelo autorregresivo no lineal. Su estado es una combinación de las entradas y salidas del patrón previo, lo que las hace idóneas para la predicción de series de tiempo. Además

de incorporar las entradas anteriores, se añaden las salidas anteriores de la red.

En esta investigación se ha entrenado una RNR NAR para realizar la predicción de la serie de tiempo sintética proveniente de la RNA MLP en un periodo de ocho años a futuro. Dicho entrenamiento se realizó en MATLAB con el módulo de *Time Series App*, que se abre ejecutando el comando *nstart* en la línea de comandos. Las características de la red (Figura 4) son las siguientes:

- Tipo RNA : recurrente NAR
- Algoritmo de entrenamiento : propagación inversa
- Algoritmo combinado : regularización bayesiana
- Estructura de la RNA : 12 neuronas
- Retardo : 96 valores

El retardo de 96 valores (ocho años) se ha establecido con base en la cantidad de años que se desean proyectar.

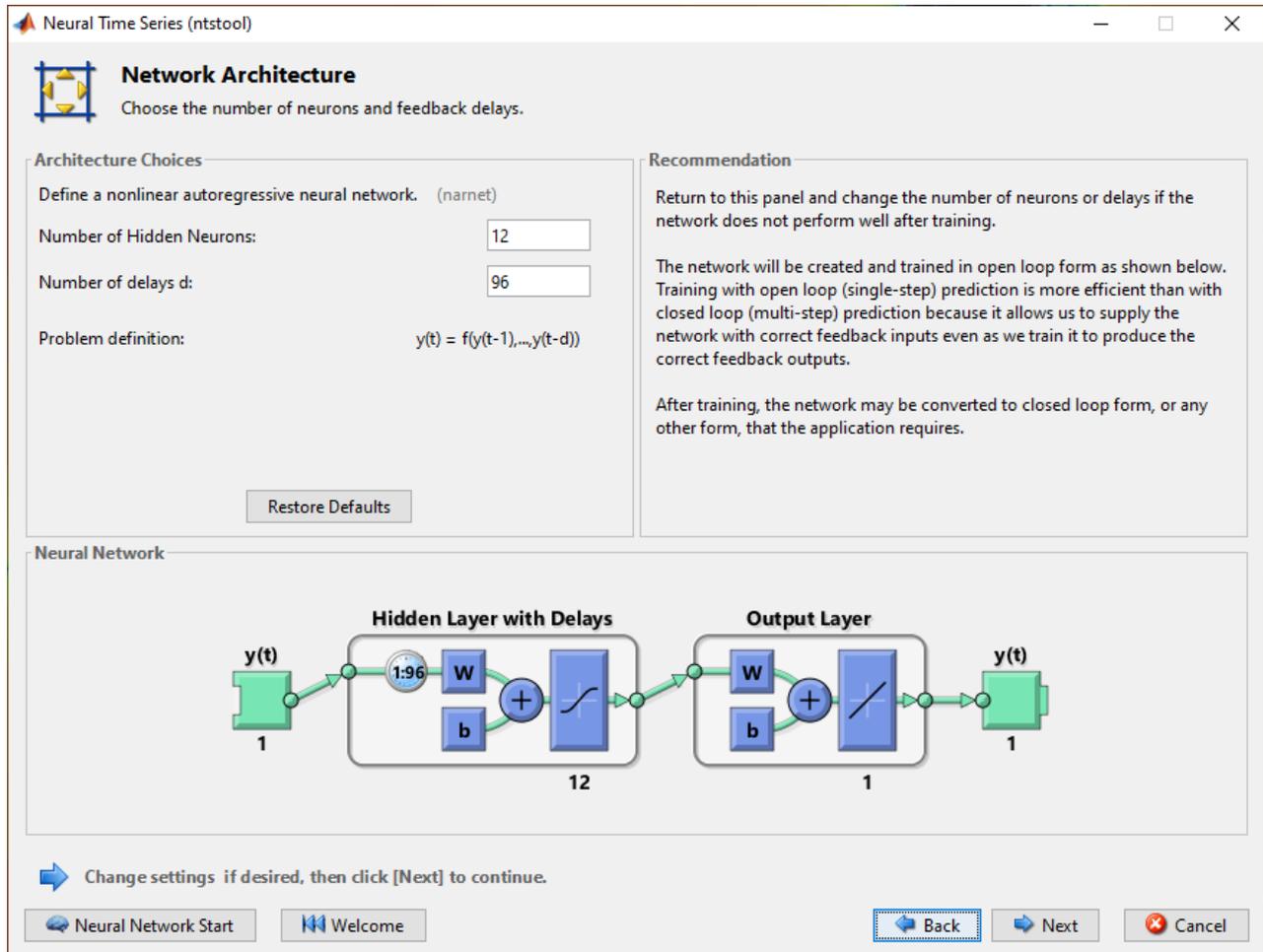


Figura 4. Entrenamiento de la RNR NAR en MATLAB 2015.

Con la red entrenada, se ejecutan las siguientes líneas de código mediante un *script* de MATLAB, que permite propagar o realizar el pronóstico a partir de la información entrenada por la RNR NAR:

```
T = tonndata(CAUDAL,false,false); %corrige la información para la red  
[x1,xio,aio,t] = preparets(net,{},{},T); %prepara la información para el tipo de red  
[y1,xfo,afo] = net(x1,xio,aio); %propaga la información a través de la red  
[netc,xic,aic] = closeloop(net,xfo,afo); %genera una red cerrada a partir de la anterior  
[y2,xfc,afc] = netc(cell(0,96),xic,aic); %realiza la propagación 96 meses u ocho años
```

Procesamiento de datos

Protocolo para implementación de RNA en modelos precipitación-escorrentía

Dawson y Wilby (2001) proponen un protocolo para la implementación de RNA en modelos de lluvia-escorrentía, el cual consiste en los siguientes pasos:

- 1° recopilar datos.
- 2° seleccionar el modelo de predicción.

3° preprocesamiento de datos-etapa 1: eliminación de saltos y tendencias; de ser necesario, remover también la estacionalidad. Seleccionar las variables a predecir y las variables que harán la predicción, elegir las más influyentes.

4° elegir un tipo de RNA: tipo de red, algoritmo de entrenamiento.

5° preprocesamiento de datos-etapa 2: escalar los datos según el rango de salida de la función de activación elegida. Para este paso puede emplearse la Ecuación (5):

$$Z_T = \frac{(Ls-Li) \times Y + (Li \cdot Mz - Ls \cdot mz)}{Mz - mz} \quad (5)$$

Donde:

Z_T = es la serie escalada.

Mz, mz = son el máximo y mínimo valor de la serie Y , respectivamente.

Ls, Li = son los límites superior e inferior a adoptar, respectivamente.

Y = es el valor a escalar.

6° entrenar la RNA.

7° validar la RNA.

Se ha escalado la información preprocesada en la primera etapa. Los parámetros requeridos para escalar cada una de las variables hacia el rango de trabajo de la función tangente hiperbólica (-1 a 1) se muestran en la Tabla 1. No se ha utilizado todo el rango de la función por la recomendación del protocolo antes mencionado, sino que se han escalado los valores de tal manera que se tenga un máximo de 0.9 y un mínimo de -0.9 en cada variable.

Tabla 1. Parámetros para escalar las variables al rango de trabajo de la función de activación tangente hiperbólica.

Variable	Estación	Mz	mz	Ls	Li
Precipitación	Cachachi	445.80	0.00	0.90	-0.90
	Cajabamba	329.50	0.00	0.90	-0.90
	Encañada	333.21	0.00	0.90	-0.90
	G. Porcón	568.90	0.00	0.90	-0.90
	Huamachuco	333.80	0.00	0.90	-0.90
	Jesús	292.24	0.00	0.90	-0.90
	L. Huangacocha	471.00	0.00	0.90	-0.90
	Namora	309.80	0.00	0.90	-0.90
	San Juan	461.80	0.00	0.90	-0.90
	San Marcos	283.30	0.00	0.90	-0.90

	S. Matara	430.20	0.00	0.90	-0.90
	A. Weberbauer	257.00	0.00	0.90	-0.90
Temperatura	Cajabamba	18.20	15.10	0.90	-0.90
	Huamachucho	15.50	11.00	0.90	-0.90
	San Marcos	20.65	14.90	0.90	-0.90
	A. Weberbauer	18.50	14.20	0.90	-0.90
	San Juan	19.50	14.80	0.90	-0.90
Cobertura	NDVI	0.52	0.31	0.90	-0.90
Caudal	Puente Crisnejas	205.60	0.62	0.90	-0.90

Recopilación y tratamiento de información meteorológica e hidrométrica

Las estaciones meteorológicas se encuentran distribuidas de manera no uniforme dentro de la cuenca y sus alrededores. Se han seleccionado las que están mejor distribuidas espacialmente en latitud, longitud y elevación, y que además poseen registros confiables en periodos de tiempo largos. La información se recopiló de las estaciones

mostradas en la Tabla 4. Las estaciones están ubicadas tal como muestra la Figura 6.

Tabla 2. Estaciones hidrometeorológicas.

Estación	Coordenadas		Altitud	Periodo de registro	Años de registro	Variables		
	Longitud	Latitud	msnm			PP	T°	Q
Augusto Weberbauer	78° 29'	7° 09'	2 660	1965-2017	53	x	x	
Cachachi	78° 16'	7° 27'	3 200	1965-2017	53	x		
Cajabamba	78° 03'	7° 37'	2 550	1965-2017	50	x	x	
La Encañada	78° 19'	7° 07'	2 980	1998-2017	20	x		
Granja Porcón	78° 37'	7° 02'	3 180	1965-2017	49	x		
Huamachuco	78° 03'	7° 49'	3 150	1965-1990 y 1991-2017	52	x	x	
Jesús	78° 23'	7° 14'	2 640	1994-2017	24	x		
Laguna Huangacocha	78° 04'	7° 56'	3 780	1965-2017	47	x		
Namora	78° 20'	7° 12'	2 760	1965-2017	53	x		
San Marcos	78° 10'	7° 19'	2 290	1965-2017	53	x	x	
Sondor Matara	78° 14'	7° 13'	2 930	1993-2017	25	x		
San Juan	78° 29'	7° 17'	2 228	1965-2017	53	x	x	

Puente Crisnejas	78° 6'	7° 27'	1,988	1968-1976 y 2014-2017	13			x
------------------	--------	--------	-------	--------------------------	----	--	--	---

Para ejecutar la primera etapa del preprocesamiento de datos propuesto por Dawson y Wilby (2001) se han filtrado los valores atípicos mediante los gráficos de cajas de Tukey (1977), y se verificaron y corrigieron los saltos en la media mediante herramientas de estadística no paramétrica, como la prueba de desviaciones acumuladas de Buishand (1982), así como en la varianza mediante la prueba de Fligner-Killeen (Fligner & Killeen, 1976). También se analizaron las tendencias con la prueba de Kendall (1975). Todo el proceso anterior se hizo en lenguaje R 3.4.0, con los paquetes Trend y Climtrends. El relleno de datos se realizó con el modelo HEC-4 del US Army Corps of Engineers (1971), que se basa en regresiones múltiples entre cada mes de registro y entre estaciones. Esta primera etapa se llevó a cabo siguiendo el flujograma mostrado en la Figura 5.

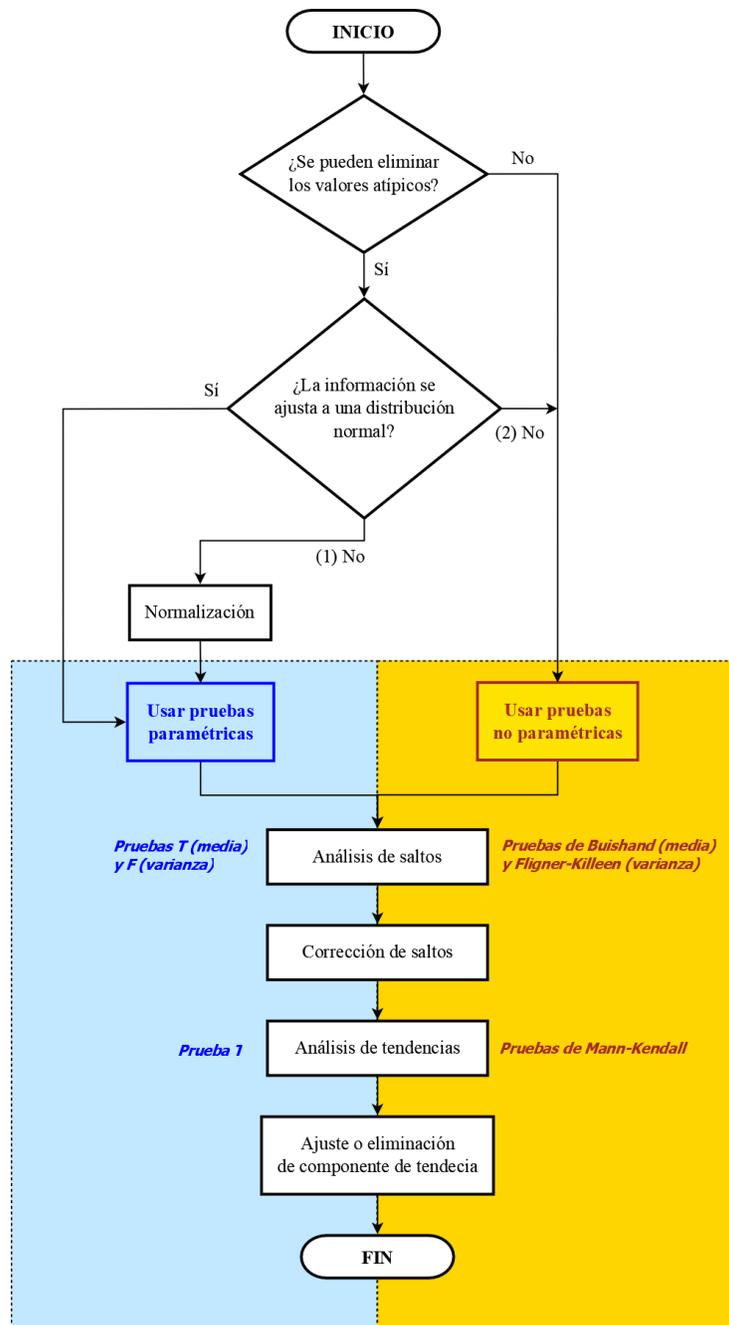


Figura 5. Diagrama de flujo de preprocesamiento de información hidrometeorológica.

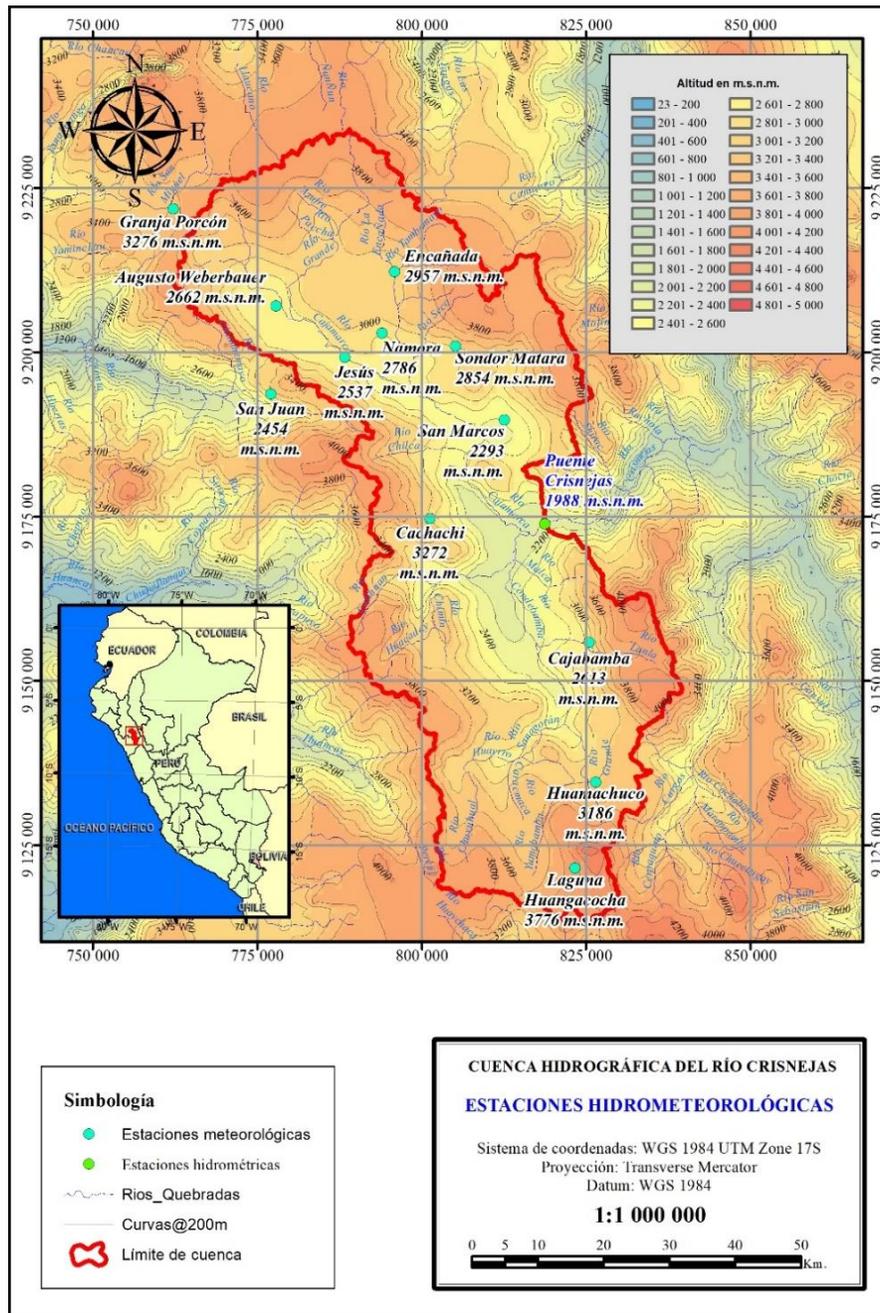


Figura 6. Estaciones hidrométricas y meteorológicas.

Recopilación y tratamiento de información cartográfica

La cuenca se ha delimitado utilizando un modelo digital de elevaciones ASTER-GDEM. Además, se adquirieron 12 imágenes multiespectrales del programa Landsat correspondientes a cada mes del año hidrológico, tal como lo muestra la Tabla 3. Estas imágenes han servido para determinar el NDVI mediante la ecuación de Huete y Tucker (1991):

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (6)$$

Donde:

NIR = es la banda correspondiente al cercano infrarrojo.

Red = es la banda del espectro rojo.

Tabla 3. Imágenes Landsat utilizadas para determinar el NDVI.

Imagen	Año de toma	Satélite	Cobertura de nubes
Enero	1992	Landsat 4-5	Menor a 10%
Febrero	1992	Landsat 4-5	Menor a 10%
Marzo	2016	Landsat 8	Menor a 30%
Abril	1995	Landsat 4-5	Menor a 10%
Mayo	1995	Landsat 4-5	Menor a 10%
Junio	2011	Landsat 4-5	Menor a 10%
Julio	2005	Landsat 4-5	Menor a 10%
Agosto	2007	Landsat 4-5	Menor a 10%
Septiembre	1984	Landsat 4-5	Menor a 10%
Octubre	1986	Landsat 4-5	Menor a 10%
Noviembre	1998	Landsat 4-5	Menor a 10%
Diciembre	1991	Landsat 4-5	Menor a 10%

Previamente al cálculo del NDVI se hicieron las correcciones y transformación de niveles digitales a parámetros físicos de cada imagen, siguiendo el flujograma de la Figura 7, adaptado de Chuvieco (1996). El procesamiento se realizó en QGIS 2.18, con uso del complemento *Semi-Automatic Classification Plugin*.

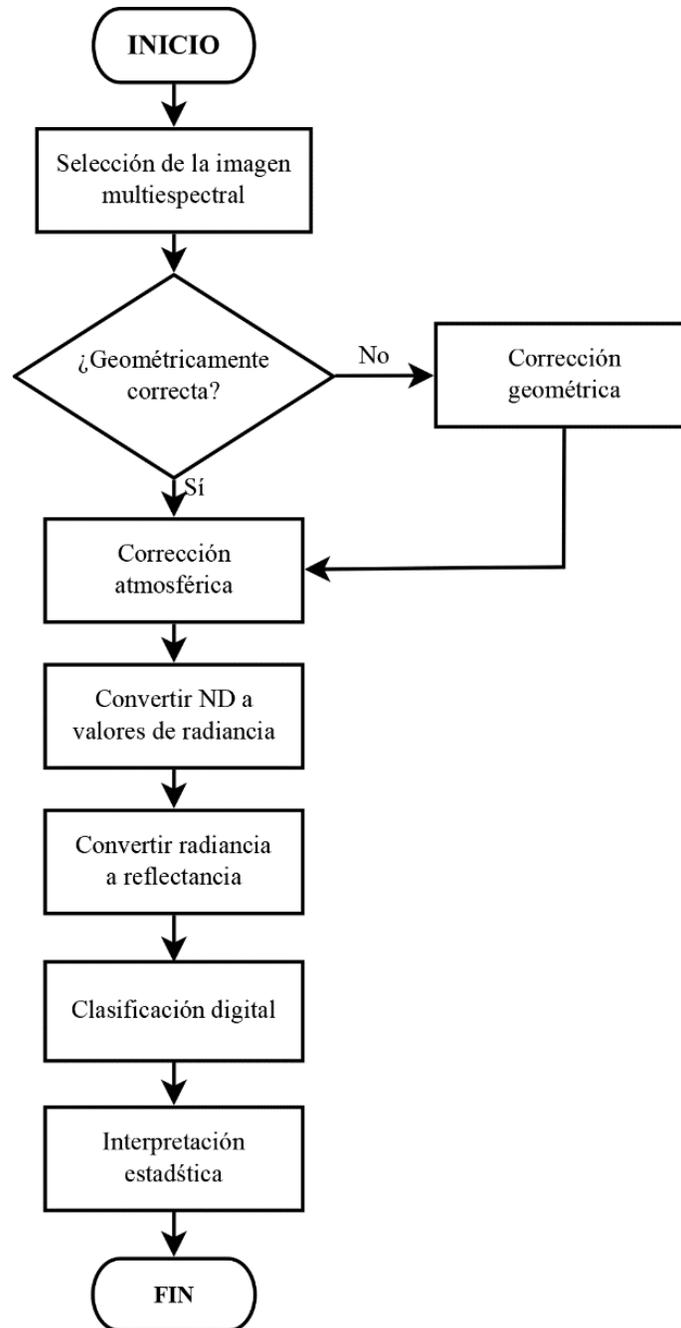


Figura 7. Diagrama de flujo de procesamiento de imágenes multispectrales.

Área de estudio

La cuenca del río Crisnejas (ver la Figura 8) se ubica en el norte del Perú, en los departamentos de Cajamarca y La Libertad. La delimitación se hizo a partir del punto ubicado en el puente Crisnejas (Tabla 4), donde está instalada una estación hidrométrica que realizó el registro de los niveles del río durante más de 30 años, pero cuyas curvas de altura-gasto no están disponibles para transformar dicha información en caudales. Solo se tienen 13 años de mediciones diarias del caudal.

Tabla 4. Ubicación de la estación hidrométrica Puente Crisnejas.

Punto	Ubicación				
	UTM-WGS 1984 Zona 17S		GCS WGS 1984		Elevación
	Este	Norte	Latitud	Longitud	
Puente Crisnejas	818705	9173905	7° 27' 48.73"	78° 6' 47.25"	1 988

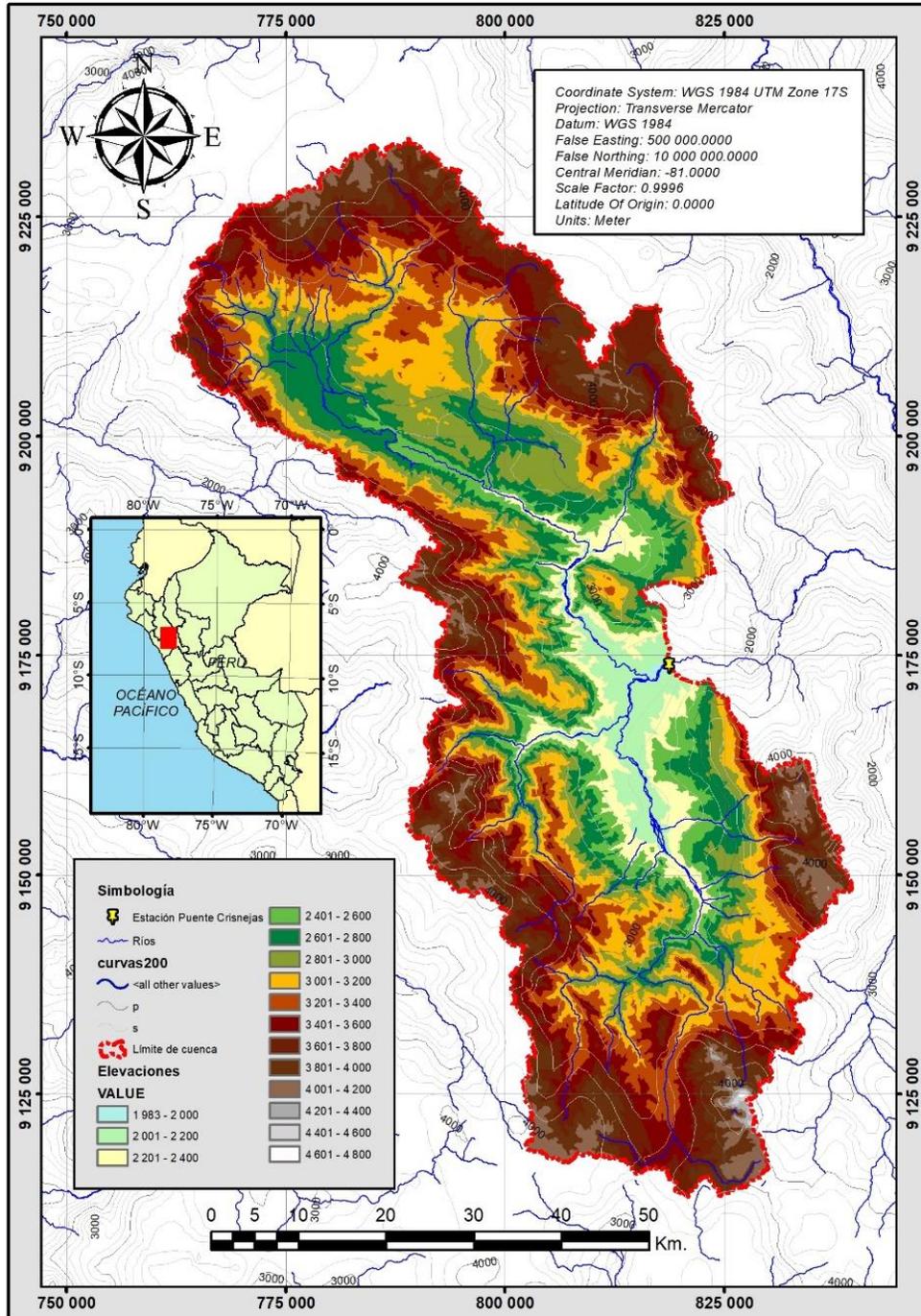


Figura 8. Cuenca del río Crisnejas.

Resultados

Análisis de la información cartográfica

Además de delimitar la cuenca se han determinado los índices de vegetación de diferencia normalizada (NDVI) para cada mes de un año hidrológico asumido como promedio, tal como lo muestra la Tabla 5 y la Figura 9. En algunos meses, la cobertura de nubes no permitió la obtención del NDVI en algunas zonas de la cuenca; no obstante, dado que el dato numérico requerido es un promedio, no se completó la información y se obtuvo solamente el promedio de lo capturado en la imagen. La Figura 10 muestra la distribución espacial del NDVI.

Tabla 5. NDVI, promedio mensual para entrenamiento de RNA MLP.

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
NDVI	0.52	0.48	0.50	0.47	0.44	0.39	0.32	0.32	0.32	0.31	0.36	0.34

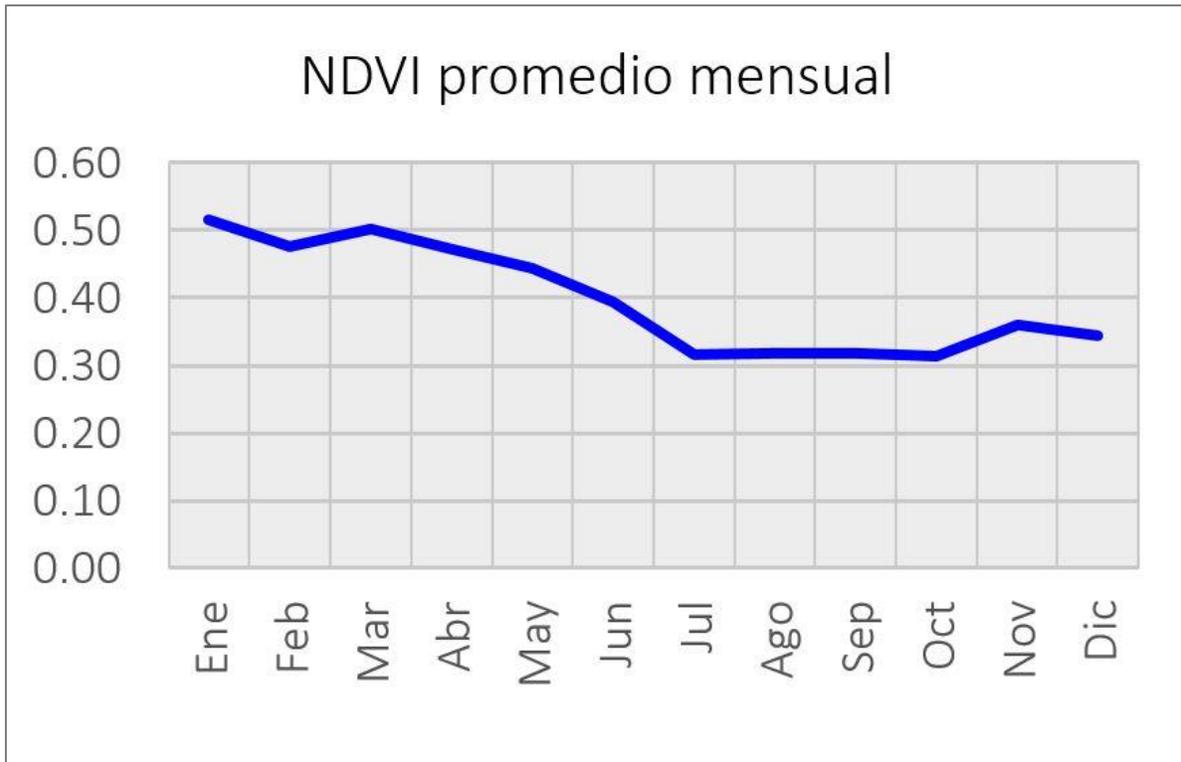


Figura 9. Índice de Vegetación de Diferencia Normalizada-NDVI, promedio mensual.

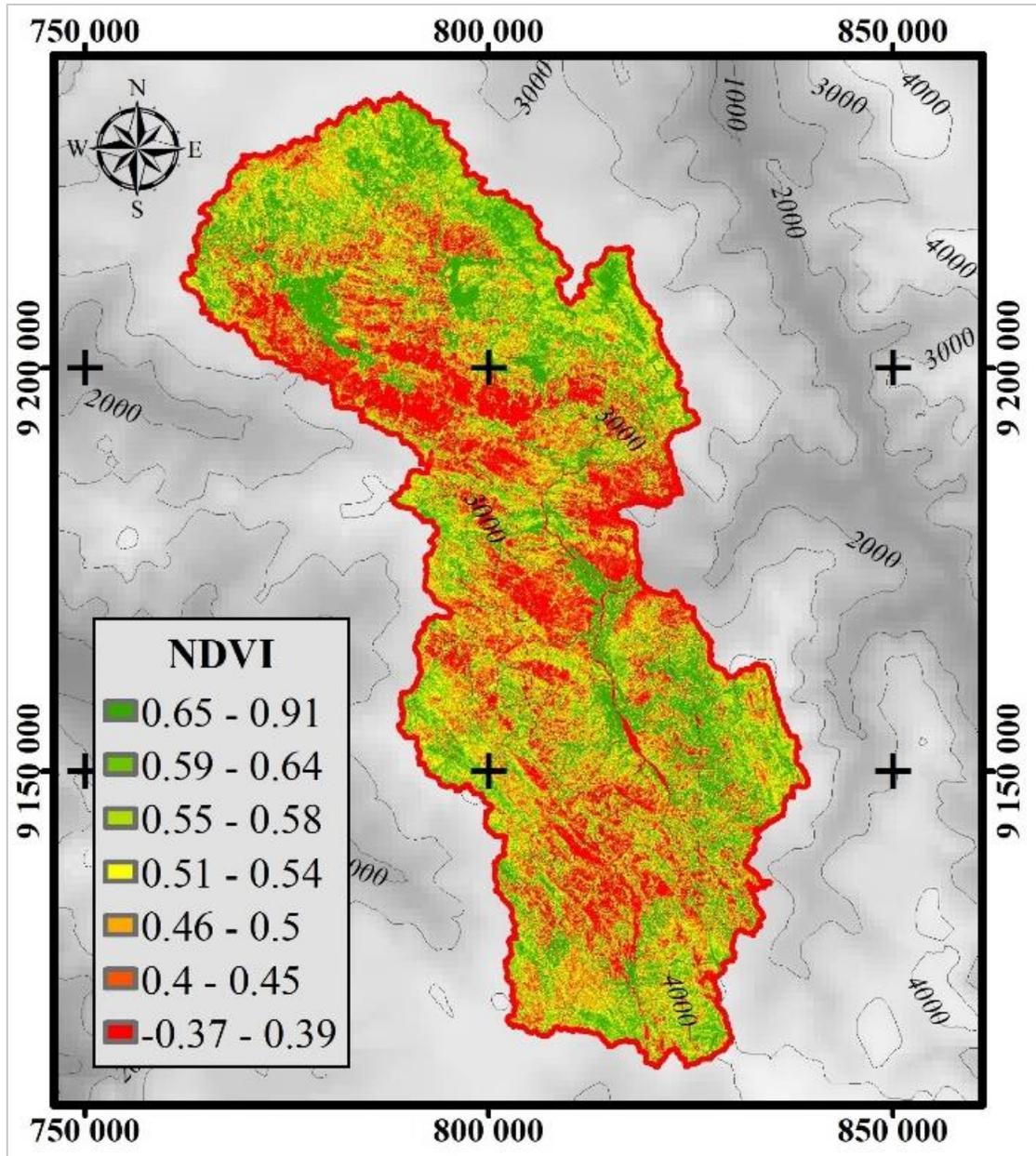


Figura 10. Cálculo del NDVI.

Análisis de la información hidrometeorológica

El procesamiento de la información hidrometeorológica resultó en la obtención de series de tiempo de precipitación y temperatura mensual homogéneas tanto en la media como en la varianza, y libres de tendencias y valores atípicos. Además, se uniformizó el registro de todas las estaciones meteorológicas mediante la extensión de serie de tiempo de registro corto (**iError! No se encuentra el origen de la referencia.** y Figura 14).

En general, el comportamiento del ciclo hidrológico de la región muestra una época húmeda en los meses de septiembre a marzo y la época seca en los meses de abril a agosto, tal como lo muestra la Figura 11. Las temperaturas señalan un comportamiento con valores más altos en la época húmeda y menores en la época seca, a excepción de la estación San Juan, donde ocurre lo inverso (ver la **iError! No se encuentra el origen de la referencia.**). Inicialmente se pensó descartar esta estación, sin embargo no se eliminó dado que su comportamiento podría enriquecer el comportamiento de la RNA MLP, y de no ser así, son los pesos en el entrenamiento los que descartan su influencia.

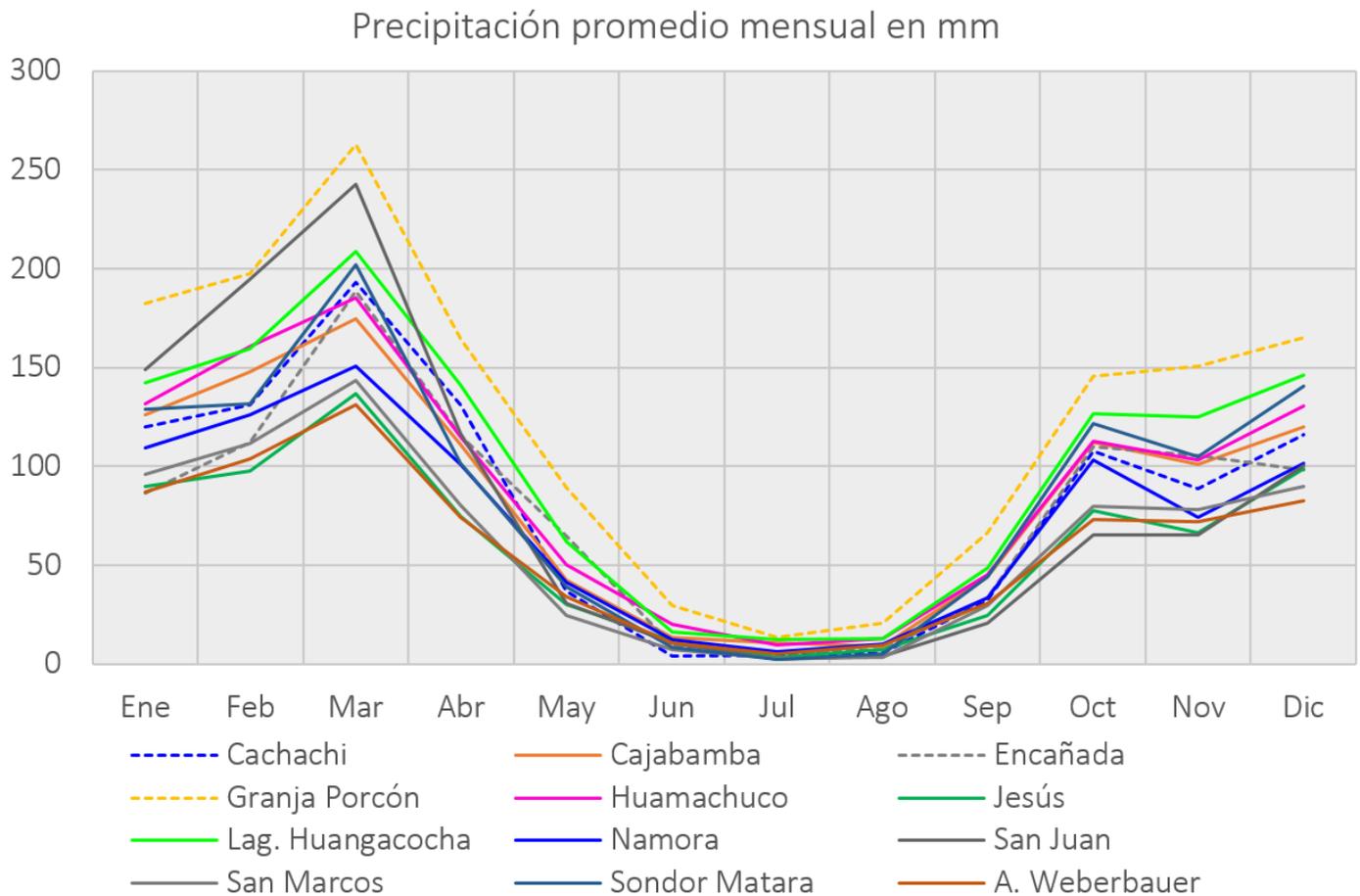


Figura 11. Precipitación media mensual en mm para registro completo y extendido en el periodo 1965-2017.

Temperatura promedio mensual en °C

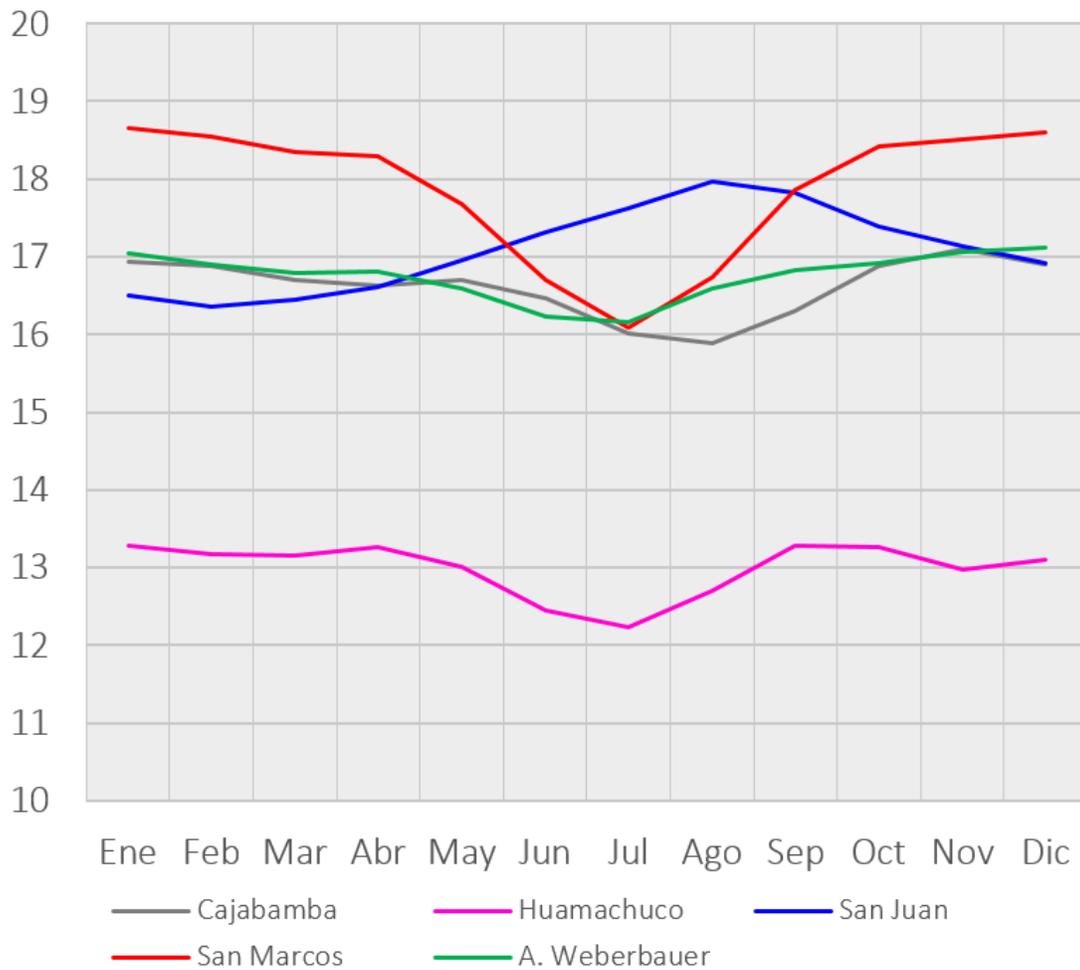


Figura 12. Temperatura media mensual en °C para registro del periodo 1965-2017.

Precipitación total anual en mm, 1965 - 2017

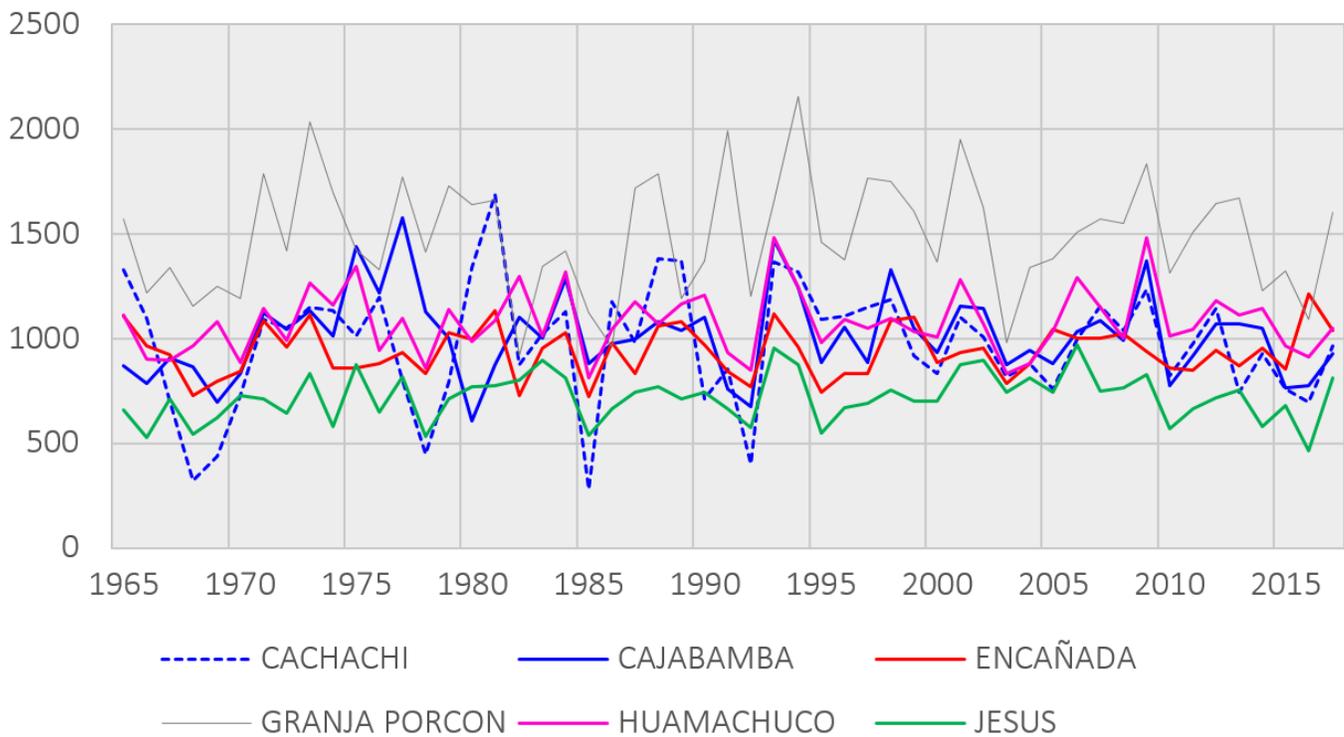


Figura 13. Precipitación total anual en mm, periodo 1965-2017.

Precipitación total anual en mm, 1965 - 2017

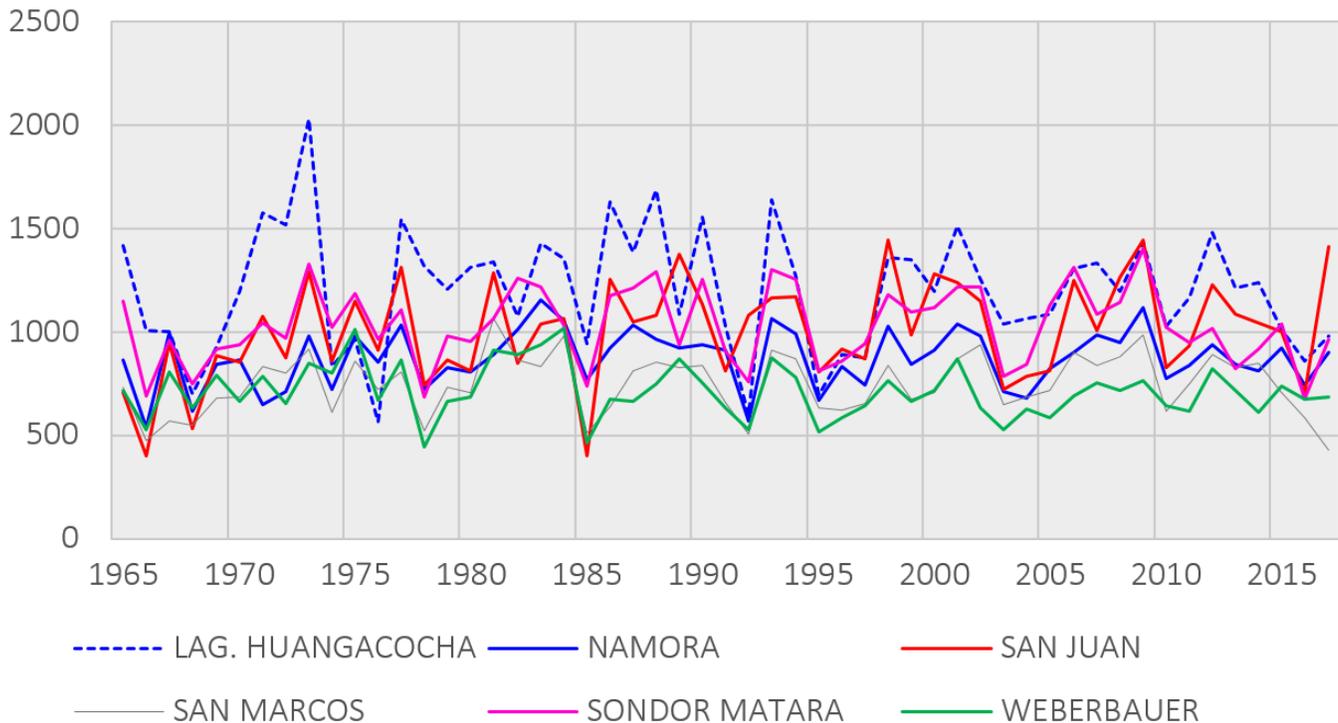


Figura 14. Precipitación total anual en mm, periodo 1965-2017.

La aplicación de herramientas de estadística no paramétrica ha permitido que el análisis de las series de tiempo sea más confiable y coherente con lo esperado del comportamiento hidrológico de la región estudiada.

El mismo procedimiento se ha seguido para el análisis de caudales de la estación hidrométrica Puente Crisnejas (Figura 15).

Homogeneización de caudales

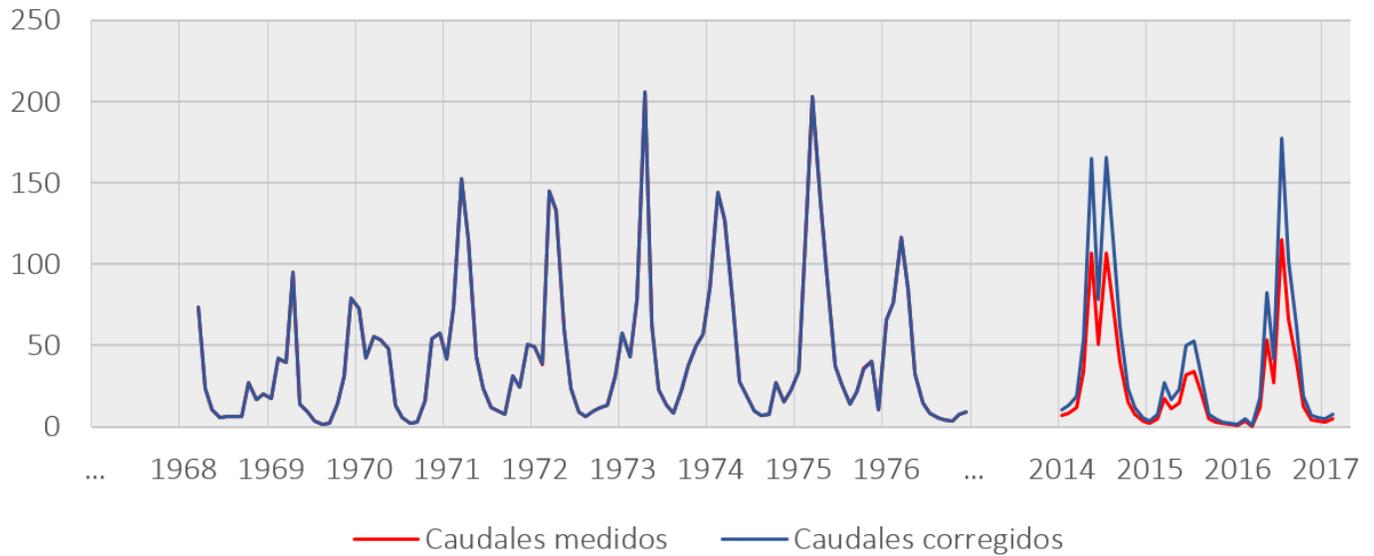


Figura 15. Homogeneización de caudales (m^3/s) en la estación Puente Crisnejas.

Entrenamiento del perceptrón multicapa (RNA MLP). Estimación de caudales en registro histórico de 1965 a 2017

El entrenamiento de la RNA MLP muestra un alto nivel de ajuste entre los datos medidos y los datos entrenados por la red, como puede verse en la **¡Error! No se encuentra el origen de la referencia..** Se hizo la validación entre los datos reservados del entrenamiento para evaluar la capacidad predictiva de la red ante los patrones no entrenados. Como es de esperarse, los datos de patrones entrenados (Figura 16) presentan, en general, un mejor ajuste con lo medido que la información generada a partir de patrones no entrenados (Figura 17); aun así, dicha información muestra un alto grado de bondad de ajuste según las medidas o coeficientes consideradas por Cabrera (2012).

Tabla 6. Bondad de ajuste de caudales estimados por la RNA tipo MLP.

Medidas de bondad de ajuste ¹	Periodo entrenado (1968-1976, 2016)		Periodo no entrenado (2014, 2015, 2017)	
	Valor	Calificación	Valor	Calificación
Coeficiente de calibración (r)	0.99	Correlación positiva fuerte	0.90	Correlación positiva fuerte
Coeficiente de determinación (r²)	0.97		0.81	
Criterio de Schultz (D)	1.38	Muy bueno	8.65	Bueno
Desviación media acumulada (MAD)	6.25	-	18.92	-
Eficiencia de Nash-Sutcliffe (E)	0.96	Excelente	0.77	Muy bueno
Error de balance de masas (m) en %	11.29	-	-3.08	-
Raíz error cuadrático medio (RMSE)	8.69	-	28.55	-

¹Cabrera (2012).

Validación - RNA MLP, patrones entrenados

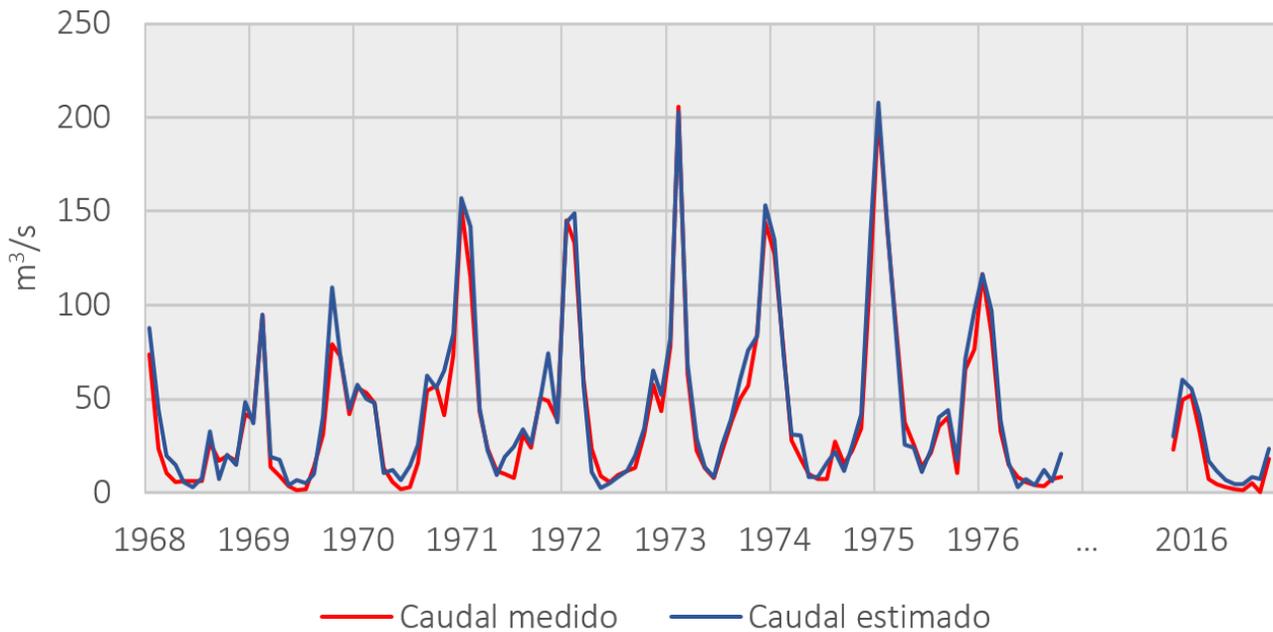


Figura 16. Aprendizaje. RNA MLP para patrones mensuales entrenados.

Validación - RNA MLP, patrones no entrenados

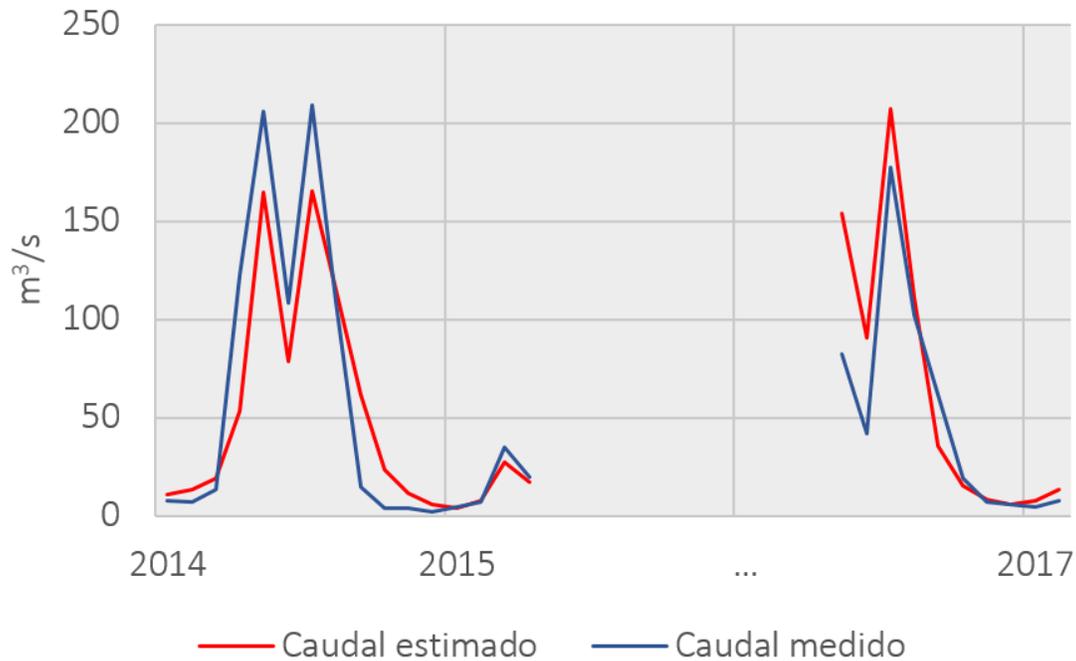


Figura 17. Validación. RNA MLP para patrones mensuales no entrenados.

Como se ve en las pendientes de las rectas de regresión, la Figura 18 indica que hay un buen ajuste entre la información medida y la estimada con la RNA MLP tanto para patrones entrenados como no entrenados.

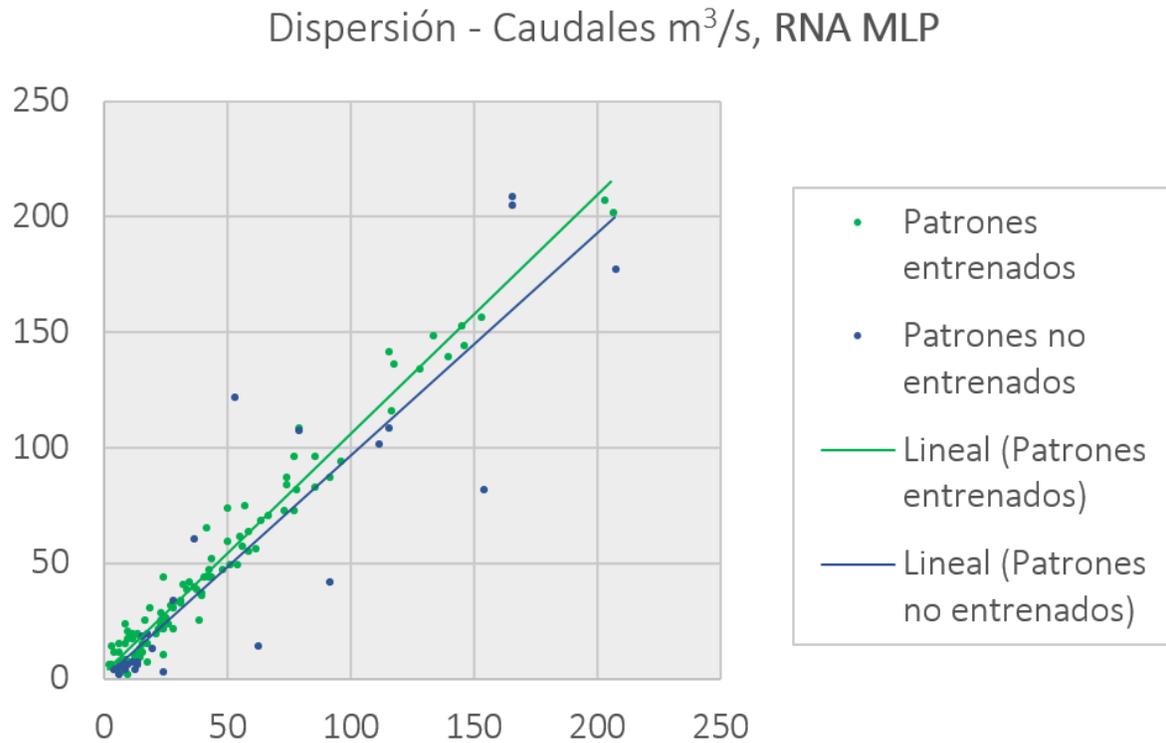


Figura 18. Dispersión de caudales mensuales, RNA MLP.

Al respecto de la valoración de la desviación media acumulada (MAD), es importante aclarar que se busca que este parámetro sea lo más cercano a 0 posible, pues representa el promedio de las diferencias entre los datos observados y los estimados. Directamente puede interpretarse el valor de 6.35 de los datos del periodo entrenado y los 18.92 del no entrenado como el “error promedio” en m³/s entre la información estimada y la medida en dichos periodos.

El error de balance de masas (m) representa, en cantidad, la relación entre el volumen del hidrograma observado y el simulado; de igual manera, tiene mejor valoración cuanto más se acerca a 0. En este caso hay menor error en los datos generados para el periodo no entrenado.

La raíz del error cuadrático medio (RMSE) cuantifica la magnitud de la desviación entre los valores medidos y los estimados; asimismo, un valor más cercano a 0 implica un mejor ajuste. Para este caso en particular, nuevamente el periodo entrenado presenta mejor ajuste que el no entrenado.

El registro de caudales mensuales generados con la RNA MLP para el periodo comprendido entre los años 1965 y 2017 se muestra en la Figura 19.

Caudal estimado (m^3/s), 1965-2017

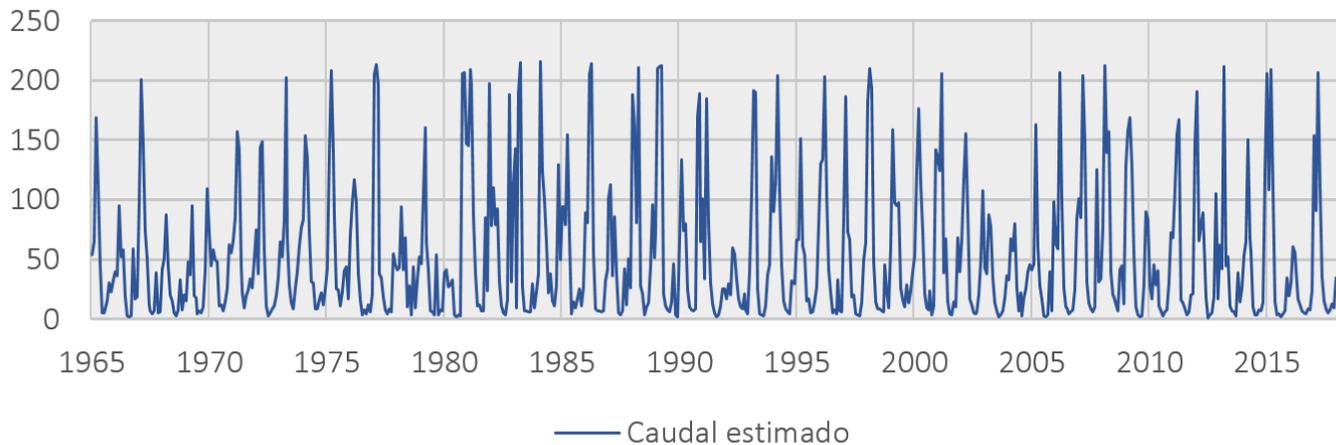


Figura 19. Registro de caudales mensuales (m^3/s) estimados con la RNA MLP, periodo 1965-2017.

Entrenamiento de la red recurrente (RNR NAR). Pronóstico de caudales en registro de 2018 a 2025

Los datos de caudal mensual estimados con la RNA MLP se utilizaron (en su forma escalada) para entrenar la RNR NAR, que dio como salida los datos de caudal mensual proyectados hasta el año 2025, tal como lo muestra la Figura 23.

El pronóstico pasó por un análisis de saltos y tendencias antes de tomarse como válido, teniendo que descartar la respuesta dada por varias redes entrenadas. Finalmente se seleccionaron los datos que no tuvieron necesidad de pasar por correcciones de este tipo.

Cabe mencionar que, dada la naturaleza de la red y el ruido en los datos, suelen generarse algunos valores negativos que resultan absurdos en el pronóstico. En este caso fueron 6 valores de los 96 meses; sin embargo, fueron purgados y reemplazados por el valor positivo inmediatamente superior. Para dar como válido el pronóstico, la red tuvo que ser entrenada muchas veces hasta conseguir que aprenda casi a la perfección el comportamiento de la serie de tiempo de caudales mensuales para reducir errores en el pronóstico. Los resultados del entrenamiento generados en MATLAB se muestran en la Figura 20 y la Figura 21; el valor de la correlación en el periodo de entrenamiento es casi perfecto. La validación hecha por MATLAB presenta también una correlación positiva moderada. Se reservaron datos mensuales de los últimos dos años para determinar la precisión del pronóstico. La **iError!** **No se encuentra el origen de la referencia.** muestra la bondad de ajuste.

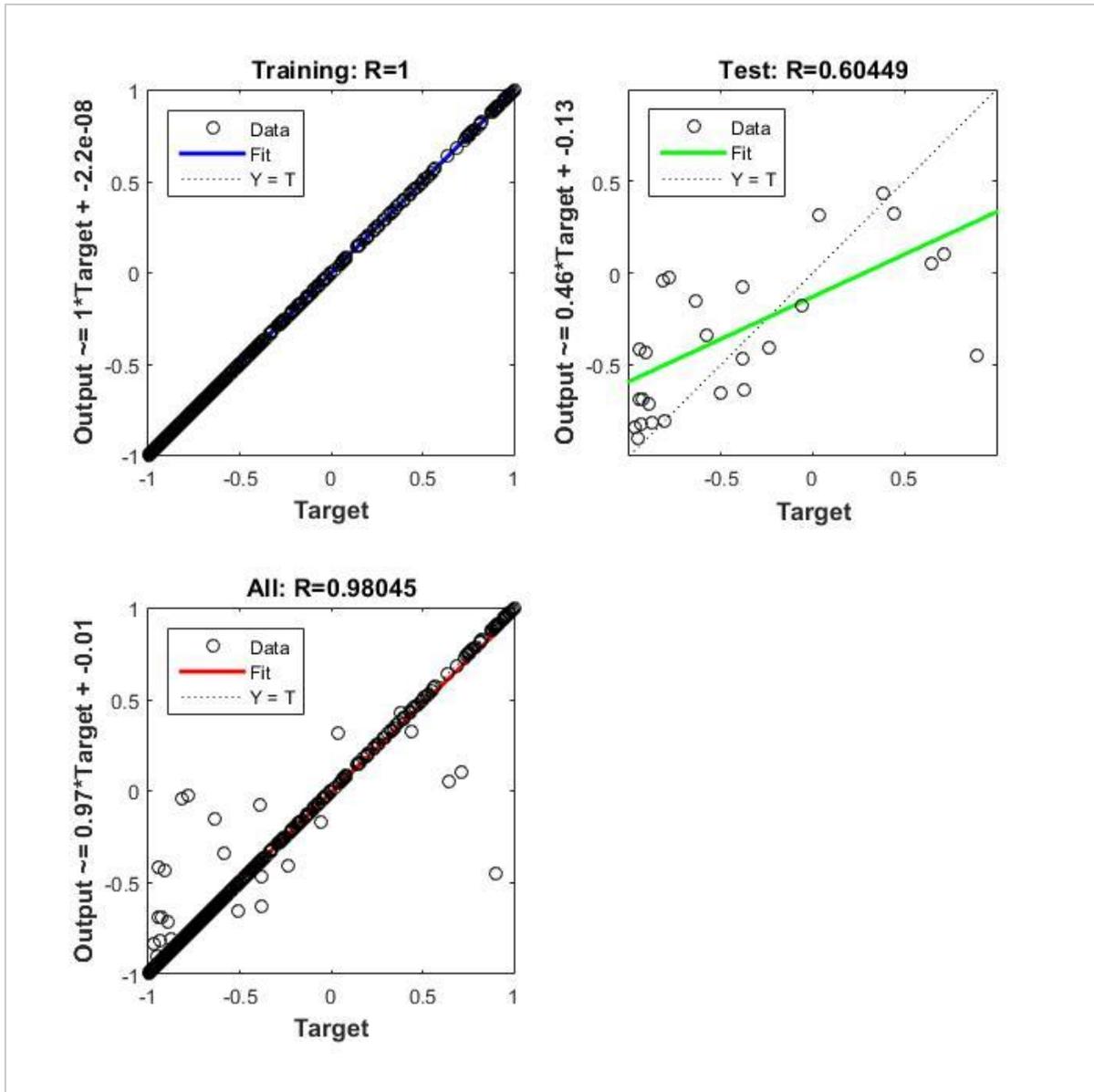


Figura 20. Coeficientes de correlación determinados por MATLAB.

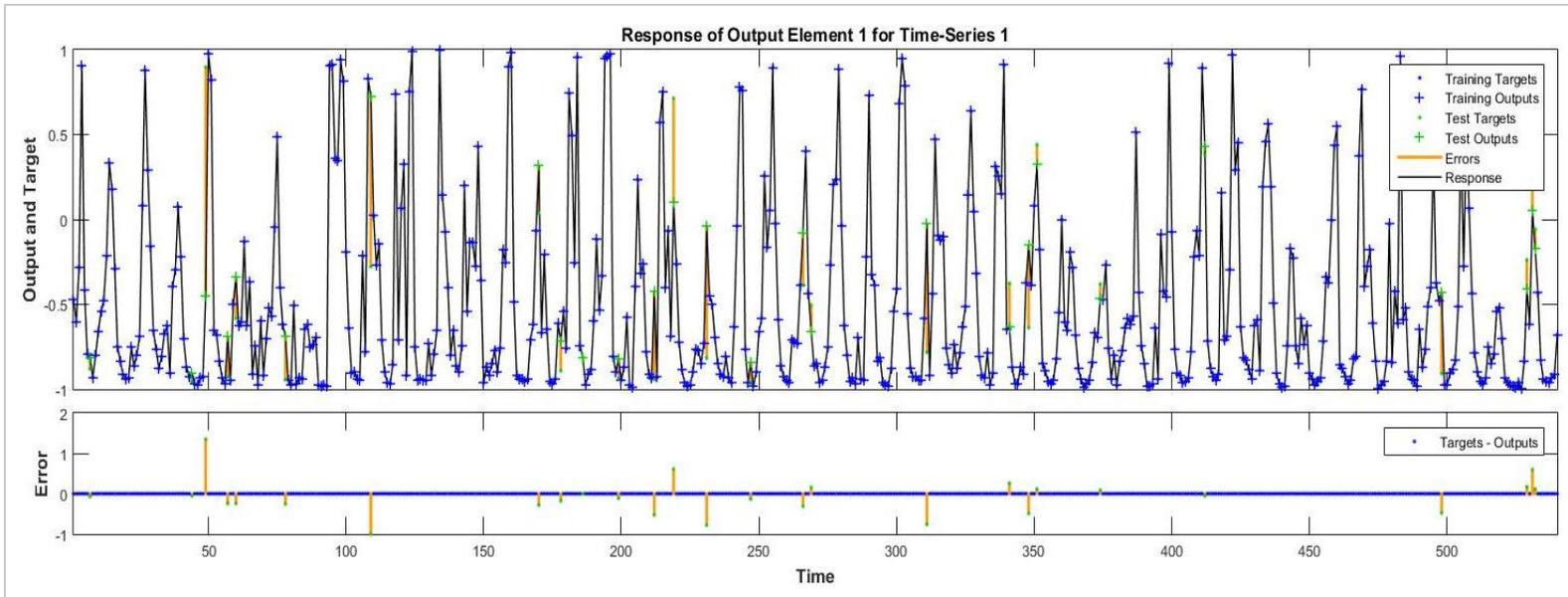


Figura 21. Respuesta de la RNR NAR para la serie de tiempo entrenada.

Tabla 7. Bondad de ajuste de caudales estimados por la RNA tipo NAR; comparativo de periodos medido y pronóstico 2017-2019.

Medidas de bondad de ajuste	Pronóstico de caudales (2018-2025)	
	Valor	Ajuste
Coeficiente de calibración (r)	0.84	Positiva-Fuerte
Coeficiente de determinación (r^2)	0.71	
Criterio de Schultz (D)	8.25	Bueno

Desviación media acumulada (MAD)	13.61	
Eficiencia de Nash-Sutcliffe (E)	0.64	Muy bueno
Error de balance de masas (m)	33.55	
Raíz del error cuadrático medio (RMSE)	23.55	

En general, los datos presentan un ajuste aceptable, teniendo en cuenta que son pronósticos, y su valor siempre puede verse afectado por variables no controladas en la simulación (crecimiento de la demanda o cambio climático) y entrenamiento de la RNR NAR.

El pronóstico podría resultar menos certero cuanto más se aleja del último dato medido, dado que el error en el pronóstico se hace más grande a cada paso de la propagación, teniendo en cuenta que cada dato generado depende de los últimos 96, lo cual sustenta la razón por la que no es conveniente el pronóstico de un largo periodo de tiempo con este tipo de técnica.

Los datos del pronóstico se muestran en la Figura 22 y la serie completa se muestra en la Figura 23.

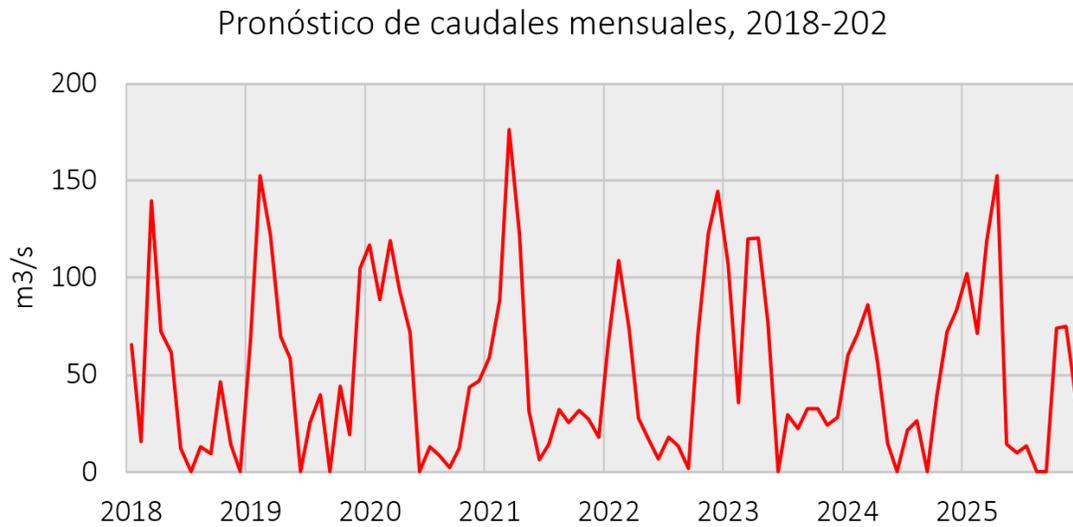


Figura 22. Pronóstico de caudales mensuales, periodo 2018-2025.

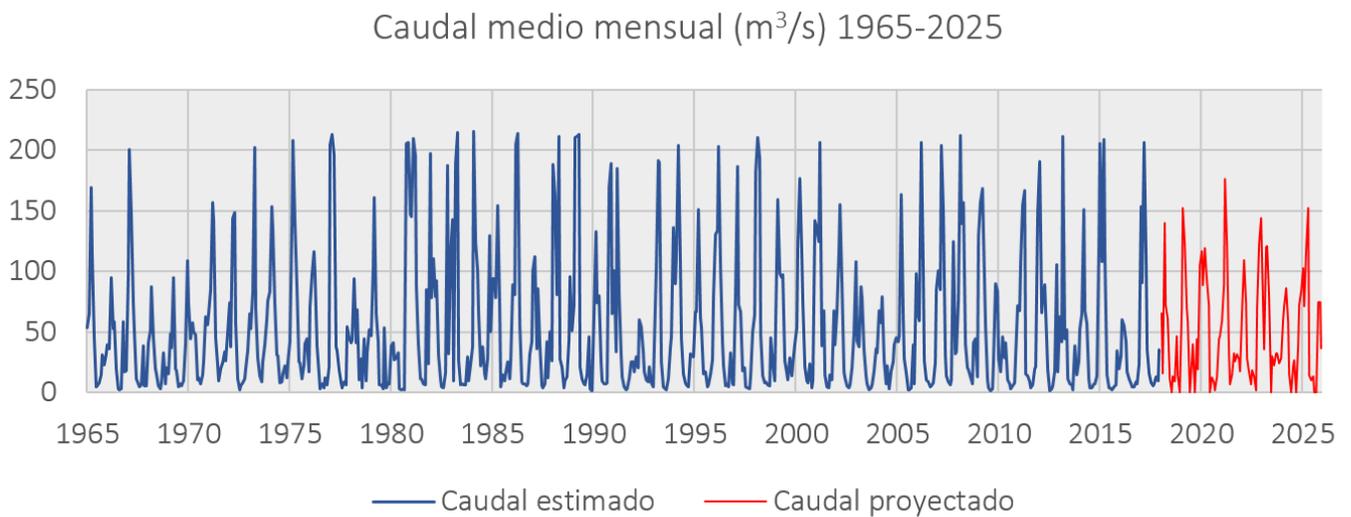


Figura 23. Pronóstico del caudal mensual en la estación Puente Crisnejas extendido hasta el año 2025.

El promedio anual de caudales se muestra en la Figura 24, con la finalidad de observar un resumen del comportamiento pronosticado por la RNR NAR.

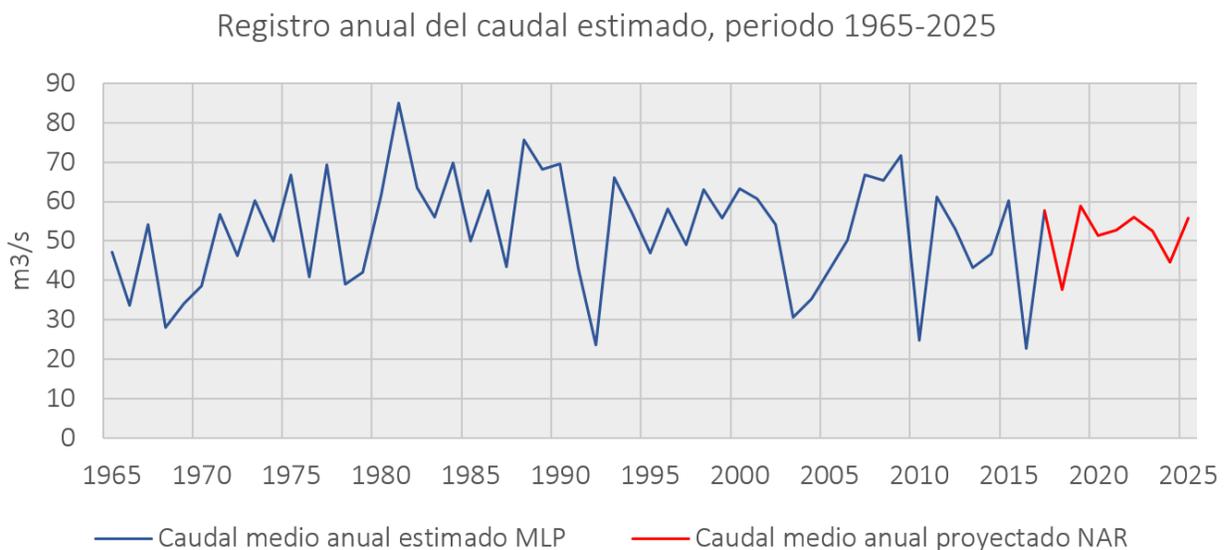


Figura 24. Registro estimado del caudal medio anual, periodo 1965-2025.

Discusión

Las redes neuronales artificiales utilizadas en la investigación han permitido la estimación del registro de caudales mensuales faltantes y el pronóstico de dichos caudales, dando como resultado una serie sintética total de 61 años de registro (1965-2025). Este registro brinda un mejor panorama de la oferta hídrica del río para la planificación y el desarrollo de futuros proyectos de aprovechamiento hidráulico.

El registro completo no presenta una tendencia considerable en los datos, sin embargo la serie de pronóstico muestra valores bajos de caudal. Esto podría deberse a los errores en la medición de los años iniciales con los cuales se entrenó el perceptrón multicapa; aun cuando la información fue corregida de saltos, la variación es notoria entre la información medida en el periodo 1968-1976 con el periodo 2014-2019. Lamentablemente es un factor que no se puede controlar dada la inexistencia de metadatos en la estación hidrométrica.

A pesar de lo anterior, los resultados de esta investigación demuestran la robustez de las redes neuronales artificiales (RNA) del tipo perceptrón multicapa y recurrente en la generación de series sintéticas de caudales mensuales a partir de información meteorológica con una alta bondad de ajuste. A su vez, se muestra una base procedimental adaptable para su extrapolación en cuencas con similar registro de información e

incluso para casos en los que se requiera una mejor resolución temporal, como la diaria, resultado que compatibiliza con los encontrados por Lama y Sánchez (2020), quienes evaluaron el efecto de las técnicas de descomposición para utilizarlas con una red neuronal recurrente llamada memoria larga a corto plazo para aumentar la precisión de la predicción diaria del caudal del río Chira en el norte de Perú; igualmente, Lee, Lee y Yoon (2019), y Heras y Matovelle (2021) obtuvieron resultados de la predicción que mostraron un buen rendimiento con errores cuadráticos medios mínimos con coeficientes de correlación altos, lo cual asegura que los modelos con RNA son adecuados para evaluar sistemas hidrológicos e hidrogeológicos complejos de aguas.

La técnica empleada ha permitido utilizar la mayor cantidad de información medida y disponible de la cuenca sin tener que recurrir a simplificaciones preliminares en las variables (estimación de otras variables mediante ecuaciones empíricas) del ciclo hidrológico y dando como resultado un registro completo con una alta bondad de ajuste.

El uso de herramientas de estadística no paramétrica ha permitido simplificar el análisis de información. No se ha tenido que recurrir a normalizaciones u otras técnicas que den validez a los datos para ser aplicados con pruebas de estadística tradicional. Es importante tener en cuenta que se ha trabajado con una cantidad relativamente grande de datos, y que en futuras investigaciones o trabajos que requieran una mejor resolución temporal, la cantidad de información a preprocesar antes del entrenamiento podría resultar muy compleja si no se considera este aspecto.

Otros trabajos de investigación en RNA para generación de series sintéticas de caudales mensuales, como el de Laqui (2010), demuestran que un esquema basado en una RNA MLP con entradas de precipitación y evapotranspiración actual y antecedente muestran mejor correlación entre lo medido y estimado que solo con los datos actuales de precipitación y evapotranspiración para su caso. Esto no necesariamente es determinante en todo tipo de cuencas teniendo en cuenta el retardo de cada una u otros factores que pudiesen influir en el comportamiento hidrológico mensual. Como en este caso no se ha investigado ninguna otra conceptualización de la cuenca en cuanto a sus variables, el entrenamiento se ha ido mejorando, modificando los parámetros de configuración de la red, como el número de capas o neuronas, e incluso la función de activación, y se ha obtenido un coeficiente de correlación aún más alto. Gomes-Villa-Trinidad (2016) aplica las redes neuronales en el pronóstico de caudal del mes siguiente utilizando una RNA MLP; sin embargo, dado que su objetivo es diferente al de esta investigación, también la conceptualización de los patrones de entrenamiento es distinta. En su caso, utiliza el caudal del mes anterior histórico y las precipitaciones y temperaturas, logrando resultados bastante buenos en la predicción del caudal para el mes siguiente, aunque como se dijo anteriormente, no se busca la generación de una serie sintética sino más bien un pronóstico. El pronóstico para la presente investigación se hizo de manera histórica y con otro tipo de arquitectura de red (RNR NAR), dado que este registro sintético permite una visualización a largo plazo útil en la toma de decisiones.

El esquema de la RNA MLP entrenado en esta investigación es un buen punto de partida para futuras investigaciones en las que se requiera la generación de series sintéticas de caudales mensuales; es importante señalar en este punto que, a diferencia de otras investigaciones, aquí no se han realizado transformaciones entre variables medidas, las variables tomadas en campo han sido las que entrenaron la RNA MLP, lo cual demuestra la ventaja de las redes neuronales artificiales en cuanto al aprovechamiento de la mayor cantidad de información medida en la cuenca.

Como puede notarse en los resultados, las redes neuronales y la información satelital tienen una amplia gama de aplicaciones en la estimación de registros y pronóstico de caudales a corto o largo plazos; solamente son limitadas por entendimiento del investigador y la adecuada selección de variables en el proceso que se requiere modelar. Herrera *et al.* (2020), al igual que nosotros, en su investigación proponen modelos basados en redes neuronales artificiales e información satelital para el llenado de datos faltantes en las estaciones meteorológicas y reconstrucción espacial de las variables de precipitación y temperatura para la región de Departamento de Valle del Cauca, Colombia, con resultados obtenidos que alcanzan los coeficientes de correlación de alrededor de 0.9.

Futuras investigaciones también podrían analizar los pesos entrenados y determinar la influencia de las mediciones de campo en cada estación con respecto al caudal e incluso tratar de interpretar el comportamiento a través de ecuaciones regionales.

Conclusiones

La generación de la serie histórica y de pronóstico de caudales mediante entrenamiento de redes neuronales artificiales ha resultado satisfactoria y con una alta bondad de ajuste, lo cual permite tener una base sólida en cuanto a la toma de decisiones en futuros proyectos de aprovechamiento hídrico de la cuenca.

Este trabajo muestra la robustez de la técnica empleada, y su alta capacidad de adaptación y aprovechamiento de la información medida en la cuenca. Se ha mostrado un protocolo adaptable a cuencas con similares registros hidrometeorológicos, como gran cantidad de cuencas en la costa y sierra peruanas, que de otra manera llevaría a recurrir a modelos de precipitación-escorrentía que no siempre dan resultados al menos aceptables, o que requieren de largos procesos de calibración o mediciones en campo adicionales respecto de los parámetros requeridos por cada modelo.

Además, se presenta un esquema y configuración de RNA MLP y RNR NAR para poder usarse como punto de partida en análisis similares.

La metodología empleada es extrapolable a muchos casos, dado que se han utilizado técnicas para el análisis, corrección y procesamiento de datos meteorológicos que se caracterizan por su amplio rango de aplicación en diferentes tipos de datos, en este caso, las técnicas de estadística no paramétrica y las redes neuronales artificiales, para las cuales existen múltiples herramientas de uso libre. Además, estas técnicas dan resultados bien ajustados sin necesidad de recurrir a asunciones o hacer supuestos respecto a los datos, y tampoco se ha tenido que recurrir a procesos de calibración.

Finalmente, la información proporcionada por esta investigación muestra la factibilidad del uso de redes neuronales artificiales como técnica de estimación de series sintéticas de caudales mensuales tanto en registro histórico como en pronósticos.

Referencias

- Alipour, M. H., & Kibler, K. M. (2019). Streamflow prediction under extreme data scarcity: A step toward hydrologic process understanding within severely data-limited regions. *Hydrological Sciences Journal*, 64(9), 1038-1055.
- Allen, G. R., Pereira, L. S., Raes, D., & Smith, M. (2006). *Guías para la determinación de los requerimientos de agua de los cultivos*. Roma, Italy: The Food and Agriculture Organization (FAO).

- Althoff, D., Ribeiro, R. B., & Neiva-Rodrigues, L. (2021). Gauging the ungauged: Regionalization of flow indices at grid level. *Journal of Hydrologic Engineering*, 26(4), 04021008.
- ANA, Autoridad Nacional del Agua, 2015. Reglamento de Procedimientos Administrativos para el Otorgamiento de Derechos de Uso de Agua y de Autorización de Ejecución de Obras en Fuentes Naturales de Agua (RJ 007-2015-ANA).
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000). Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering*, 5(2), 124-137.
- Blanco, P. A. (2016). *Algoritmo de retropropagación*. Recuperado de http://www.cs.us.es/~fsancho/ficheros/IAML/2016/Sesion04/seminario_BP.pdf
- Brenes, A. (2020). *Predicción del caudal promedio horario de la estación hidrológica palmar, utilizando modelos de machine learning basados en árboles de decisión* (doctoral dissertation, tesis de grado). Universidad de Costa Rica, Costa Rica.
- Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, 58(1-2), 11-27.
- Cabrera, J. (2012). *Calibración de modelos hidrológicos*. Lima, Perú: Universidad Nacional de Ingeniería, Facultad de Ingeniería Civil, Instituto para la Mitigación de los Efectos del Fenómeno El Niño.

- Canchala, T., Alfonso-Morales, W., Carvajal-Escobar, Y., Cerón, W. L., & Caicedo-Bravo, E. (2020). Monthly rainfall anomalies forecasting for southwestern Colombia using artificial neural networks approaches. *Water*, 12(9), 2628.
- Choubin, B., Solaimani, K., Rezaeehad, F., Roshan, M. H., Malekian, A., & Shamsirband, S. (2019). Streamflow regionalization using a similarity approach in ungauged basins: Application of the geo-environmental signatures in the Karkheh River Basin, Iran. *Catena*, 182, 104128.
- Chuvieco, E. (1996). *Fundamentos de teledetección espacial* (No. 526.982 C564 1996). Madrid, España: Ediciones Rialp.
- Darras, T., Johannet, A., Vayssade, B., Kong-A-Siou, L., & Pistre, S. (2018). Ensemble model to enhance robustness of flash flood forecasting using an Artificial Neural Network: case-study on the Gardon Basin (south-eastern France). *Boletín Geológico y Minero*, 129(3), 565-578.
- Dawson, C. W., & Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, 25(1), 80-108.
- Delgado, A. (1998). *Inteligencia artificial y minirobots*. Bogotá, Colombia: Ecoe Ediciones.

- Farfán, J. F., Palacios, K., Ulloa, J., & Avilés, A. (2020). A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: Methodology and case studies in Andean watersheds. *Journal of Hydrology: Regional Studies*, 27, 100652.
- Fattorelli, S., & Fernández, P. C. (2007). *Diseño hidrológico*. Mendoza, Argentina: Zeta Editores.
- Fligner, M. A., & Killeen, T. J. (1976). Distribution-free two-sample tests for scale. *Journal of the American Statistical Association*, 71(353), 210-213.
- Gomes-Villa-Trinidad, G. (2017). *Modelo de predicción de aportaciones mensuales utilizando Redes Neuronales Artificiales. Aplicación a la cuenca del río Amambáí (Brasil)*. Recuperado de <http://hdl.handle.net/10251/78117>.
- Heras, D., & Matovelle, C. (2021). Machine-learning methods for hydrological imputation data: Analysis of the goodness of fit of the model in hydrographic systems of the Pacific-Ecuador. *Revista Ambiente & Água*, 16.
- Herrera, W. M., Leiva, O. F. B., & Romero, M. E. R. (2020). Aplicación de redes neuronales para la reconstrucción de series de tiempo de precipitación y temperatura utilizando información satelital. *Revista EIA*, 34(17), 1-16.

- Herrera-Quispe, J., Yari, Y., Luque, E., & Tupac, Y. (2013). Red neuronal aplicada a la generación de caudales mensuales estocásticos. XIII Congreso de la Sociedad Peruana de Computación-CSPCAT, Chiclayo, Perú. DOI: 10.13140/2.1.4047.7762
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., & Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrological Sciences Journal*, 58(6), 1198-1255.
- Huete, A. R., & Tucker, C. J. (1991). Investigation of soil influences in AVHRR red and near-infrared vegetation index imagery. *International Journal of Remote Sensing*, 12(6), 1223-1242.
- Isasi-Vinuela, P., & Galván-León, I. M. (2004). *Redes de neuronas artificiales. Un enfoque práctico*. Madrid, España: Editorial Pearson Educación, S. A.
- Kalteh, A. M. (2008). Rainfall-runoff modelling using artificial neural networks (ANNs): Modelling and understanding. *Caspian Journal of Environmental Sciences*, 6(1), 53-58.
- Kendall, M. (1975). *Rank correlation measures*. London, UK: Charles Griffin.

- Lama, G. L. R., & Sánchez, I. (October, 2020). Hybrid models based on mode decomposition and recurrent neural networks for streamflow forecasting in the Chira River in Peru. In: 2020 IEEE Engineering International Research Conference (EIRCON) (pp. 1-4), Institute of Electrical and Electronics Engineers, Lima, Perú.
- Laqui, W. (2010). Aplicación de redes neuronales artificiales a la modelización y previsión de caudales medios mensuales del río Huancané. *Revista Peruana Geo-Atmosférica RPGA*, 2, 30-44.
- Larranaga, P., Inza, I., & Moujahid, A. (1997). Tema 8. Redes neuronales. En: *Redes neuronales*. Lejona, España: Universidad del País Vasco, Recuperado de <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t8neuronales.pdf>
- Lee, S., Lee, K. K., & Yoon, H. (2019). Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. *Hydrogeology Journal*, 27(2), 567-579.
- Osborn, H. B., & Lane, L. (1969). Precipitation-runoff relations for very small semiarid rangeland watersheds. *Water Resources Research*, 5(2), 419-425.

- Pérez-Ortiz, J. A. (2002). *Modelos predictivos basados en redes neuronales recurrentes de tiempo discreto* (tesis doctoral). Departamento de Lenguaje y Sistemas Informáticos, Universidad de Alicante, España. Recuperado de <https://rua.ua.es/dspace/bitstream/10045/3826/1/Perez-Ortiz-Juan-Antonio.pdf>
- Razavi, T., & Coulibaly, P. (2016). Improving streamflow estimation in ungauged basins using a multi-modelling approach. *Hydrological Sciences Journal*, 61(15), 2668-2679.
- Scholz, L. (1980). *Generación de caudales mensuales en la sierra peruana*. Cusco, Perú: Plan Meriss II.
- Steyn, M. (2018). *Short-term stream flow forecasting and downstream gap infilling using machine learning techniques* (Doctoral dissertation). Stellenbosch, South Africa: Stellenbosch University.
- Témez, J. R. (1977). *Modelo matemático de transformación precipitación-aporación*. Madrid, España: Comisión E. Explotación y Garantía, Grupo de Trabajo de Predicciones de Precipitación y Relación entre Precipitaciones y Caudales.
- Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical Review*, 38(1), 55-94.
- Torres, G. H., & Granados, G. R. S. (2019). *Predicción hidrológica en cuencas no instrumentadas-estado del arte*. Encuentro Internacional de Educación en Ingeniería. Recuperado de <https://doi.org/10.26507/ponencia.273>.

Tukey, J. W. (1977). *Exploratory data analysis*. Vol. 2 (pp. 131-160).
Recuperado de http://theta.edu.pl/wp-content/uploads/2012/10/exploratorydataanalysis_tukey.pdf

USACE, US Army Corps of Engineers. (1971). *Hec-4 monthly streamflow simulation*. Washington, DC, USA: US Army Corps of Engineers.