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

Articles

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

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

Luis Vásquez-Ramírez[1], ORCID: https://orcid.org/0000-0001-8861-7889

Luis Félix Vásquez-Paredes[2], ORCID: https://orcid.org/0000-0003-1982-4483

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

[2]Consultor, Perú, lfvp505@gmail.com

Corresponding author: Luis Vásquez-Ramírez, lvasquez@unc.edu.pe

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.

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.

Received: 12/22/2020

Accepted: 09/18/2021

Introduction

Estimating the water supply is a recurring problem in hydrology when there is no adequate record of flows in the basin. For this purpose, theoretical models are available based on the interrelation of the variables of the water cycle and the processes that help determine the amount of water available at a point of interest. The information available and required in the basins determines the characteristics of the model that can be applied in each case; therefore, sometimes, simplifications or assumptions must be made regarding the variables or the hydrological cycle, depending on the information required by the model. The selection of the variables and the amount of data determines the model's predictive capacity (Cabrera, 2012).

Torres and Granados (2019) mention that traditionally, hydrological analysis has been based on the availability of climatological and hydrological information in a hydrographic basin that allows, together with the analysis of geographical, geological, and environmental conditions, the simulation of natural phenomena diverse such as drought, floods, sudden floods, availability of water supply among others that turn out to be essential inputs for the integral management of water. However, classical modeling protocols cannot be applied when instrumentation is lacking in a basin. Hydrologists face the problem of indirectly quantifying water resources, sometimes with little scientific support. Researchers such as Alipour and Kibler (2019), and Choubin *et al*. (2019) agree that the reliable estimation of the flow, especially in uncalibrated basins, is of utmost importance for environmental management and planning and the prediction of the flow in uncalibrated basins is necessary to support the decisions taken around the best use of water.

Sivapalan and Wagener, cited by Hrachowitz *et al*. (2013), indicate that at the beginning of the new millennium, a community awareness had been reached that hydrological theories, models, and empirical methods were largely inadequate for predictions in uncalibrated basins. Furthermore, there was a need to understand better links between hydrological function, that is, how a watershed responds to inputs and shape, that is, physical properties of a watershed, to address the challenge of uncalibrated watersheds adequately.

Therefore, in the last decades, there has been a need to find new methodologies capable of improving the precision of flow predictions in uncalibrated basins. Alipour and Kibler (2019) present a method for the prediction of current flow under the extreme data scarcity model (SPED), a framework designed for the prediction of current flow within regions of dispersed hydrometeorological observation, while Razavi and Coulibaly (2016), and Choubin *et al*. (2019) propose to consider the integral characteristics of the watersheds through multiple model approaches to improve the continuous estimation of the daily flow in uncalibrated watersheds through regionalization, the process of transferring hydrological data from calibrated to non-calibrated watersheds. Currently, many researchers are including digital elevation models to improve the approximation in the calculations; this is the case of Althoff, Ribeiro, and Neiva-Rodrigues (2021), who present a methodology based on the use of the terrain analysis toolset using the model elevation (TauDEM) to obtain the input variables for the regionalization model averaged for the catchment area of each pixel in the flow network grid.

Hrachowitz *et al*. (2013), after concluding their research entitled "A decade of predictions in uncalibrated basins (PUB): a review", they found that the main factors that contribute to the resulting predictive uncertainty, which were identified by the PUB initiative, include : a) An incomplete understanding of the set of processes that underlie the response of the hydrological system, and the feedbacks at the catchment scale between these processes, which frequently results in inherently unrealistic models with high predictive uncertainty; b) An incomplete understanding of the multi-scale Spatio-temporal heterogeneity of processes in different landscapes and climates, as the vast majority of small catchments around the world were, and still are, unmeasured with little or no information available; and c) Inadequate regionalization techniques to transfer understanding of hydrological response patterns from measured to unmeasured environments due to a lack of cross-basin comparative studies and a lack of understanding of the physical principles that govern sound regionalization.

In small basins, or cases in which little data are available, or specific precipitation events, the direct relationship between rainfall and runoff can be determined using regression methods (Osborn, 1969), deriving equations that can relate the flow with the rain and/or more variables (USACE, 1971). These techniques give greater flexibility in terms of the information required, although with a more significant number of assumptions and without a known interrelation between the variables involved in the process, compared to hydrological models. Furthermore, by the nature of the method, the extrapolation of values is limited, non-linear relationships cannot be solved without transforming the inputs, and it is sensitive to outliers.

In contrast to the flow estimation techniques and models described, artificial neural networks (hereinafter ANN) have advantages in that it is not necessary to know the physical relationship between the variables involved in the problem, they are robust (they do not have high sensitivity to errors in the input patterns), the input variables can be adapted to the available data (Delgado, 1998) and depending on the type of ANN, they can be applied in recurring processes to make time series forecasts. For Herrera, Leiva, and Romero (2020), in hydrology, there are many cases where neural networks have been used to predict the behavior of a variable based on previous historical data and a set of predictor variables since their research addressed the particular problem of reconstruction of missing information from meteorological stations using RNAs.

In the last decades, the use of neural networks in hydrological modeling has increased due to their fundamental property as a universal and parsimonious approximator of non-linear functions. In the field of flood forecasting, feedforward and recurrent multilayer perceptrons have confirmed their efficiency (Darras, Johannet, Vayssade, Kong-A-Siou, & Pistre, 2018). As the sustainable management of water resources requires forecasting of flows in short times, hydrological challenges that Steyn (2018) and Lama and Sánchez (2020) propose to face with the application of machine learning techniques both to treat the discontinuity of the data, as well as to work with flows that follow non-linear or stationary behaviors. While Brenes (2020) further specifies the prediction of the hourly average flow using Machine Learning models based on decision trees, comparing their predictive capacity at the Palmar hydrological station, located on the Grande de Térraba river in the South-Pacific region of Costa Rica.

For Heras and Matovelle (2021), computational methods based on machine learning have had wide development and application in hydrology, especially for modeling systems that do not have enough data. Within this problem, there are missing data series that should not necessarily be discarded; This is achieved by completing them, understanding that this requires combining approaches or methodologies. In this sense, some investigations have been developed that have had satisfactory results, such as that of Canchala, Alfonso-Morales, Carvajal-Escobar, Cerón, and Caicedo-Bravo (2020), evaluated the performance of the combination of three Artificial Neural Networks (ANN) approaches in the forecast of monthly rainfall anomalies for southwestern Colombia, or that of Farfán, Palacios, Ulloa and Avilés (2020), who propose a hybrid technique, using the time series generated by the individual models as inputs to a new ANN. This approach aims to increase the precision of the simulated flow by combining and exploiting the information provided by physical and data-driven models.

In the Crisnejas River, located in northern Peru, there is a monthly flow record of 13 years in two periods separated by a data gap of 37 years; however, complete records of precipitation and temperature are available at many weather stations in and around the basin. This situation is common in basins of the Peruvian coast and highlands of great interest in implementing hydraulic projects for which it is necessary to know water availability. The short registration period prevents an adequate probabilistic estimation of the persistence of flows, and for this reason, the registration must be completed based on rainfall-runoff relationships (ANA, 2015). In this sense, in the country, hydrological models are frequently applied for monthly flows, such as that of Témez, of global valuation and basins below 10,000 km2 (Témez, 1977) or the model developed by Lutz Scholz, within the framework of the Technical Cooperation of the Republic of Germany for the Meris II Plan, and which applies only to basins in the Peruvian highlands (Scholz, 1980).

The aforementioned hydrological models require simplifying the precipitation data from the stations into an average record within the basin, eliminating variability. The same occurs with temperature, and in the case of Témez, it is also required to estimate the average potential evapotranspiration (ETP) in the basin. Still, there is not always sufficient data, and one must opt for ETP estimation models based on temperatures. In the calibration and validation process of these models, absurd values can be found in parameters such as aquifer discharge or delay and runoff coefficients since they cannot always be applied in the basin of interest or there are simply deficiencies in the input data.

Faced with the proposal to estimate monthly flows through the aforementioned hydrological models, the ANNs do not eliminate the variability of the precipitation data from the different climatic stations but instead establish their influence on the output data implicitly or internally. Similarly, it happens with the temperature data or the additional variables that can be considered in the analysis. Furthermore, calibration is unnecessary since the ANN will seek to "learn" how it should relate the inputs to reach the output with the least possible error (Delgado, 1998). This provides a lot of flexibility regarding input information and the quality of the results.

Therefore, this research aims to apply artificial neural networks (ANNs) to estimate the missing flow data in the Crisnejas river from data on precipitation, temperature, and vegetation cover quantified by the Normalized Difference Vegetation Index. (NDVI) of an average year.

Background

Artificial neural networks are a computational technique inspired by the work of the biological neuron model and threshold logic of Warren McCulloch and Walter Pitts in 1943; the principle of the perceptron was established in 1958 with its limitation to solving only separable problems linearly, it is not until 1975 when the reverse propagation algorithm or ' backpropagation ' is proposed, and this limitation is resolved (Delgado, 1998).

Investigations that directly apply artificial neural networks (ANN) to solve complex hydrological problems have been increasingly frequent, given the large number of computational tools developed in recent years for training ANNs and their different algorithms and types. A summary of the previous works that precede this research is presented below.

The *Journal of Hydrologic Engineering* (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000) presents an article that compiles the possible applications of ANNs in various branches of hydrology, such as rain-runoff, flows, groundwater, water quality, and precipitation. It indicates that, with adequate training, ANNs can generate satisfactory results for predicting problems in hydrology.

Dawson and Wilby (2001) propose a protocol for implementing artificial neural networks in precipitation-runoff processes and flood prediction in which mention is made of a process of normalization or typification of the data in the range that is accepted by the wake-up function.

Kalteh (2008) performs a precipitation-runoff and ANN modeling using precipitation, temperature, flow, and time data. His research concluded with reasonable precision in the estimation of flow through ANNs; in addition, he points them out as promising tools not only in model precision but also in the learned relationship since he used neural interpretation methods to interpret the connection between the weights of the network.

In his research, Laqui (2010) uses the precipitation, evapotranspiration, and flow data of the Huancané River (Peru) for the training of a multilayer perceptron type ANN with the ' backpropagation ' algorithm and compares its results with a series model stochastic temporal, obtaining a better fit with the ANN.

Herrera-Quispe, Yari, Luque, and Tupac (2013 also used multilayer perceptron ANN with the Levenberg-Marquardt algorithm to generate stochastic monthly flows in the Chili River basin (Peru) in combination with the Thomas-Fiering stochastic model.

Gomes-Villa-Trinidad (2016) presents, in his master's thesis, a prediction model of monthly contributions using ANN in the Amambaí river basin (Brazil). Their conclusions showed that using ten hidden neurons could obtain better results than with networks of 15 to 25 neurons. In addition, it concludes that ANNs are a very efficient alternative to perform flow predictions in contrast to the Naive model of trivial prediction. This research also compiles the methodology proposed by Dawson and Wilby (2001) in the form of a protocol to implement precipitation-runoff models with ANN. The study also used ANN of the multilayer perceptron type with the Levenberg-Marquardt algorithm.

Materials and methods

OPEN ACCESS

Methodological proposal

To determine the historical monthly flow in the period 1965-2017 and make its prediction in eight years, the training of two artificial neural networks of the multilayer and recurrent NAR perceptron type is proposed.

For the first network, the training patterns have the following data as input:

- Precipitation. Registered monthly (1965-2017) in 12 meteorological stations in the study area (limit of the basin and surroundings).
- Temperature. They were recorded (1965-2017) by 5 of the 12 previous stations.
- Ground cover. Quantified from the NDVI and obtained from multispectral images of each month's average hydrological year.
- Flow rates. Short monthly record (1965-1976, 2014-2019), used in three parts: one for the training of the multilayer perceptron type network (1968-1976, 2016), another for the validation of said network (2014, 2015, 2017) and another shorter period (2018-

2019) for the validation of the prediction made with the recurrent type network NAR.

The diagram proposed in Figure 1 shows the process followed for training and prediction with both networks.

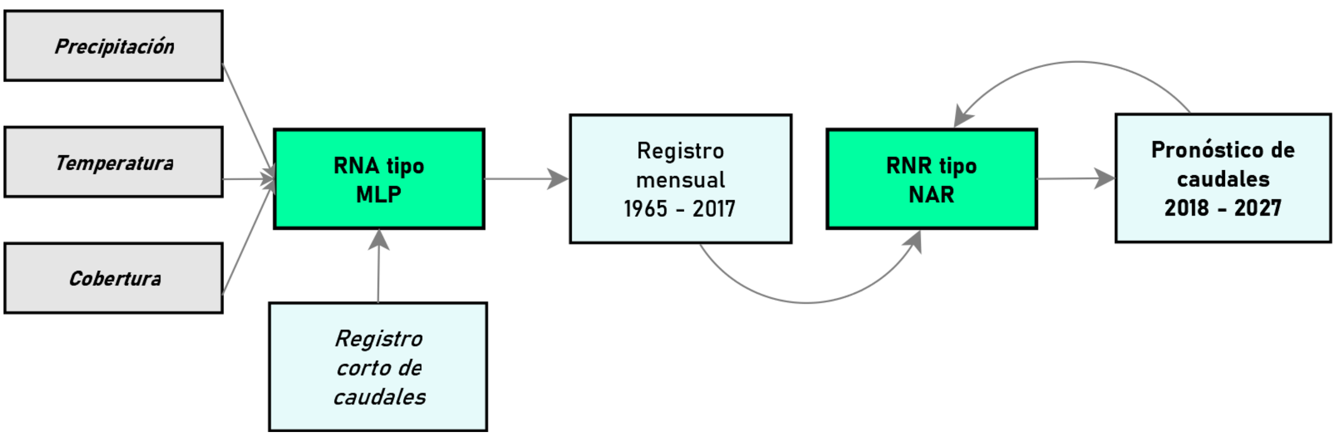

Figure 1. Methodology for estimating flows.

For the second network (RNN NAR), only the throughput data estimated with the MLP ANN is required.

Currently, there are many tools for training artificial neural networks, from programming languages such as Python or R to programs with a graphical interface such as MATLAB. For this case, the training of the MLP-type ANN has been done by encoding the backpropagation algorithm in the VB.net language. In the case of NAR-type RNN, the ' *Time Series app*' of artificial neural networks from MATLAB 2015 has been used.

Hydrological balance

The proposed methodology for estimating the monthly flow (m 3 / s) in the Crisnejas river basin is based on the approach of the most influential variables in the basin's water balance. According to Fattorelli and Fernández (2007), the hydrological model of a basin is based on the processes that integrate the phases of the hydrological cycle. In a basin, we can find several variables classified into inputs (precipitation), outputs (runoff, underground flow, evapotranspiration), and storage variation. All these variables are interrelated, as shown in Equation **¡Error! No se encuentra el origen de la referencia.**:

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

Where:

ΔS = storage in mm/year per basin area

P = precipitation in mm/year by basin area

Q = flow in mm/year by basin area

G = flow of groundwater out of the basin in mm/year per basin area

ET = evapotranspiration in mm/year by basin area

When analyzing each of the variables, it is observed that the knowledge of precipitation is essential for estimating the flow; in this case, it is considered independent of other factors and is measured data already considered in the ANN input pattern.

The underground flow depends on the cover, the type of soil, and the geology; These last two are considered constant on the monthly time scale and the global period analyzed (53 years); therefore, the parameter to be quantified coverage. In this sense, the quantification of this parameter has been proposed through the NDVI or Normalized Differential Vegetation Index according to Huete and Tucker (1991), in an average year.

Evapotranspiration, according to Allen, Pereira, Raes, and Smith (2006) is the combination of two separate processes by which water is lost through the soil surface by evaporation and transpiration of vegetation. There are many equations or methods for its estimation. In this research, its simplest conceptualization has been taken. Thornthwaite

(1948) poses Equation **¡Error! No se encuentra el origen de la referencia.**, which gives an estimate of the ETP in mm/day:

$$ETP = 16(10 \times T/I)^{a} \tag{2}$$

Where:

T = temperature in °C.

I = annual heat index, which is a function of the monthly temperature.

a = parameter as a function of I.

This way, potential evapotranspiration does not need to be entered directly into the model since it can be expressed as a function of temperature. Its behavior will also be improved from the NDVI since, in reality, it also depends on the basin's coverage.

Storage is related to complex processes in which coverage, soil type, geology, infrastructure, and relief must be considered. Its variability is not significant in the investigation's period and time scale; therefore, it is a constant.

Finally, the conceptual model is proposed to estimate the monthly flow based on precipitation (P), temperature (T), and NDVI.

$$Q = f(P, T, NDVI) \tag{3}$$

Multilayer perceptron network and backpropagation algorithm

According to Isasi-Vinuela and Galván-León (2004), unlike the simple perceptron, the multilayer perceptron allows for solving non-linearly separable problems. This type of network is composed of several hidden layers that will enable decision regions. The multilayer perceptron, or MLP (Multi-Layer Perceptron), is usually trained through the reverse propagation algorithm or backpropagation (Back Propagation), which is why the name of back propagation network also knows it.

RNAs of the multilayer perceptron type (Figure 2) are composed of an input layer, one or more intermediate or hidden layers, and an output layer. Each of the neurons in the previous layers connects with all the neurons in the following layers. The information propagates in one direction; once the information is presented in the ANN in the input layer, it reaches the output layer through the hidden ones; this process is called feedforward.

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

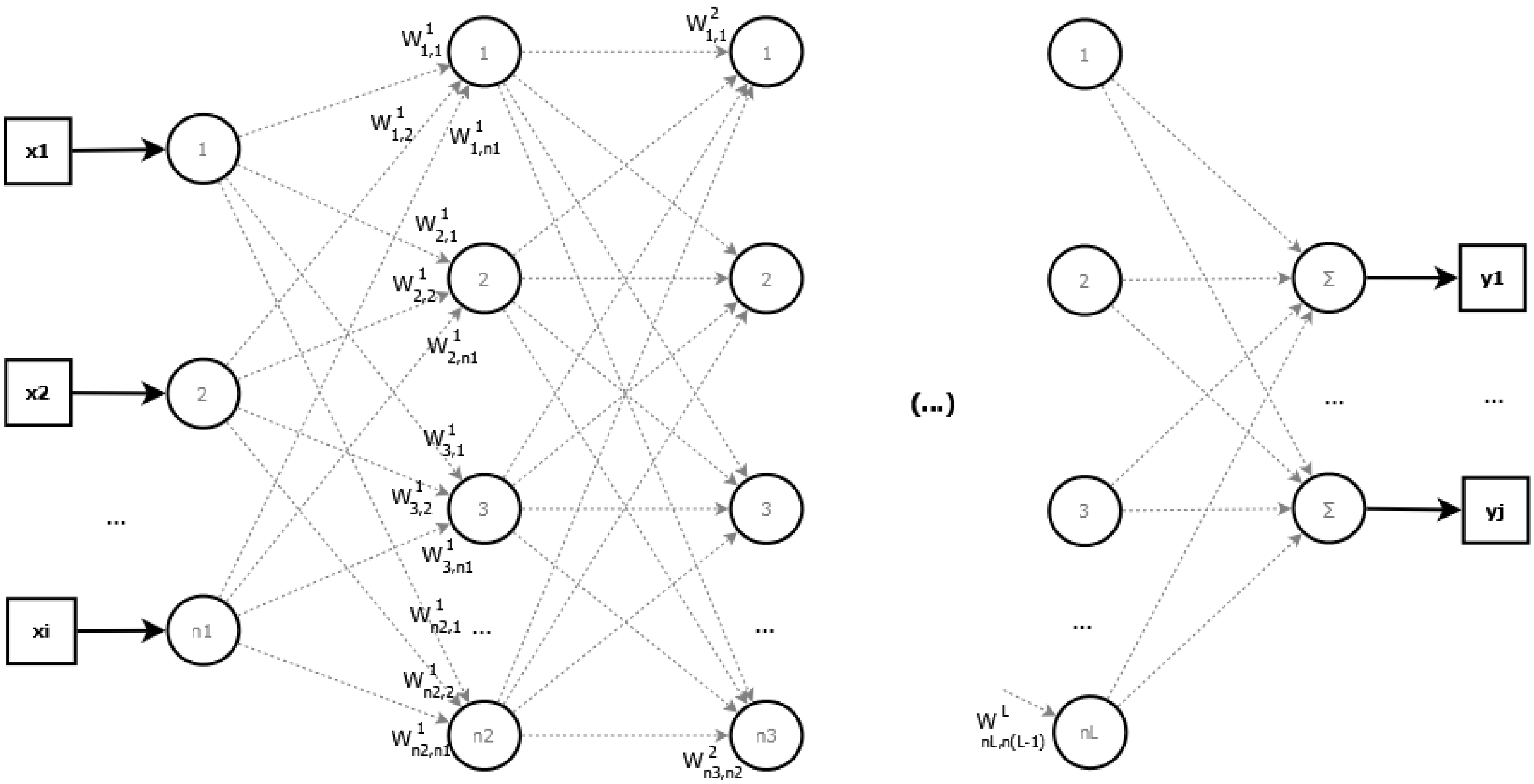

Figure 2. Multilayer perceptron.

Each neuron receives a linear combination (summation) of the information affected by the so-called "weights" and is then evaluated by the "activation function", the same one that generates the input for the next layer, as interpreted from Equation **¡Error! No se encuentra el origen de la referencia.**, according to Delgado (1998). The weights are adjusted through the training process, for which there are algorithms such as backpropagation that are combined with error minimization techniques, such as gradient descent, Levenberg-Marquardt, Newton, or conjugate gradient:

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

Where:

$y =$ neuron output.

$\sigma =$ represents the activation function (F.A.); it can be of the tangent, logistic, identity, ReLU, Gaussian, or other types.

$x =$ inputs.

$W =$ weight.

$W0 =$ activation threshold.

Backpropagation algorithm

The backpropagation algorithm to train an MLP (Multi-layer perceptron or multilayer perceptron) architecture consists of five elementary steps, according to Larranaga, Inza, and Moujahid (1997), which are:

Step 1. Randomly set initial weights and thresholds (t: = 0, initial epoch).

Step 2. For each pattern in the training set:

2.1 Execute a phase to obtain the network's response in the pattern.

2.2 Calculate the total error in the output layer.

2.3 Calculate the partial increase in weights and thresholds due to each training pattern.

Step 3. Calculate the current total increment, extended to all patterns. The same procedure is carried out with the thresholds.

Step 4. Weights and thresholds are updated

Step 5. The total error is determined, and if it is not acceptable, all the patterns are presented to the network again. The algorithm is repeated from Step 2 until satisfactory results are obtained (t: = t + 1, next epoch).

Blanco (2016) indicates that the backpropagation algorithm is usually combined with some learning algorithm such as the delta rule or the gradient descent method. With the latter, the training of the ANN used in this research has been carried out.

The proposed training scheme is shown in Figure 3. The main characteristics of the ANN used to estimate the historical record of monthly flows for the period 1965-2017 are:

- ANN Type : multilayer perceptron
- Training algorithm : reverse propagation
- Combined algorithm : gradient descent
- Unique activation function : hyperbolic tangent
- ANN structure : 7 - 5 - 4 - 1 neurons per layer

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

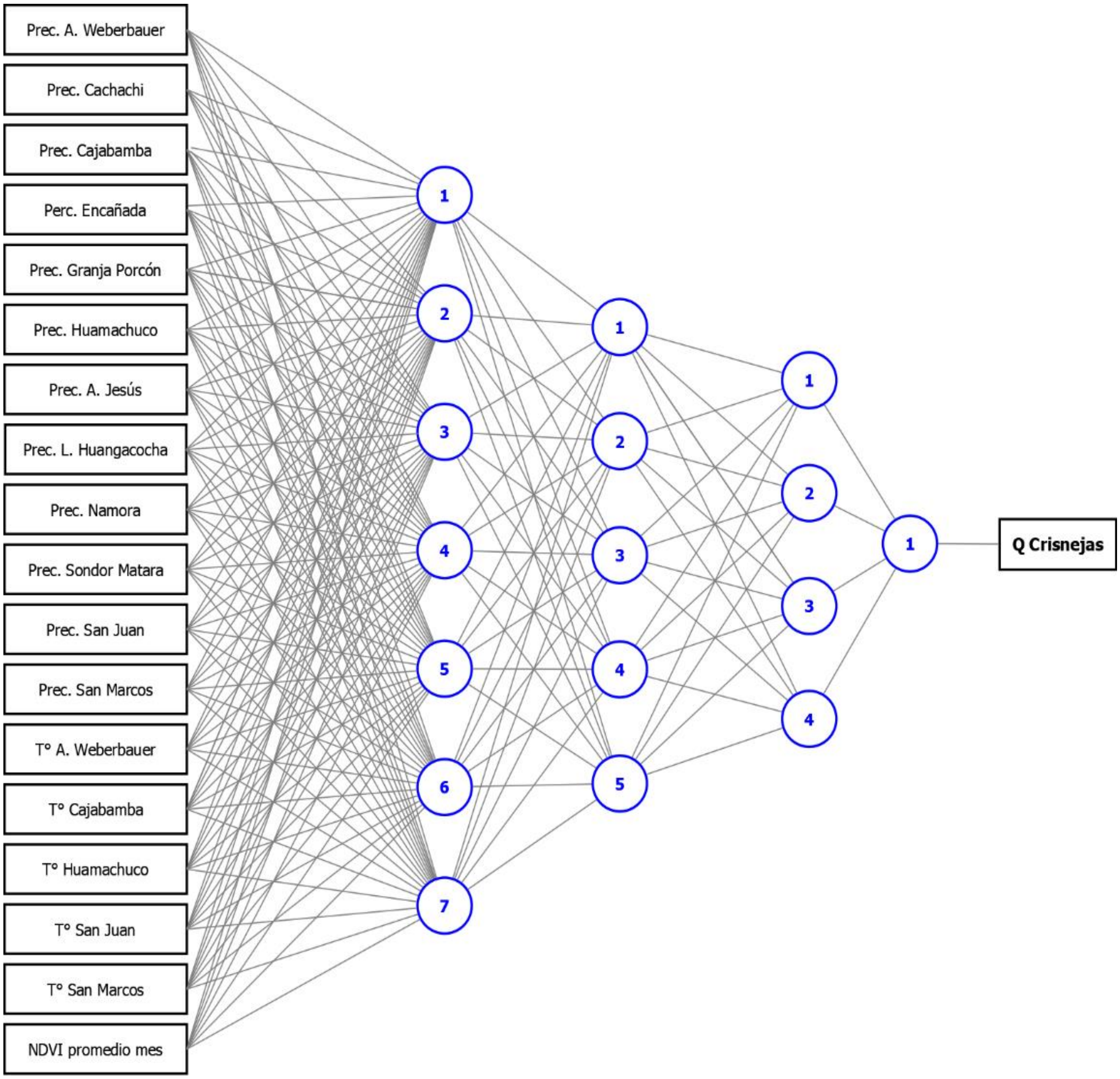

Figure 3. Trained Multilayer Perceptron Artificial Neural Network.

This scheme has been obtained from multiple trial and error processes, which acquired the best training results and the extension of untrained values.

Recurrent neural network

Pérez-Ortiz (2002) explains in his doctoral thesis that the way an ANN's neurons are interconnected defines a directed graph. If the graph is acyclic, we are dealing with the most common case of a forward-propagating or feedforward ANN, a type of network in which the multilayer perceptron-type RNAs seen above are found. In the case that the network has cycles, it is called Recurrent Neural Network. In this type of network, the existing cycles have a profound impact on the learning capacity of the network and make them more efficient for processing time series.

A recurrent neural network (RNN) can be of several types. In this research, a NAR (Nonlinear Autoregressive) type RNN is also known as a non-linear auto-regressive model. Their state is a combination of the previous pattern's inputs and outputs, making them ideal for time series prediction. In addition to incorporating the inputs above, the prior network outputs are added.

In this research, an RNN NAR has been trained to predict the synthetic time series from MLP ANN in a future period of 8 years. Said training was carried out in MATLAB with the 'Time Series app' module, which is opened by executing the 'nnstart' command on the command line. The characteristics of the network (Figure 4) are as follows:

- ANN type : Recurrent NAR
- Training algorithm : Reverse propagation
- Combined algorithm : Bayesian regularization
- ANN structure : 12 neurons
- Delay : 96 values

The delay of 96 values (eight years) has been established based on the number of years to be projected.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

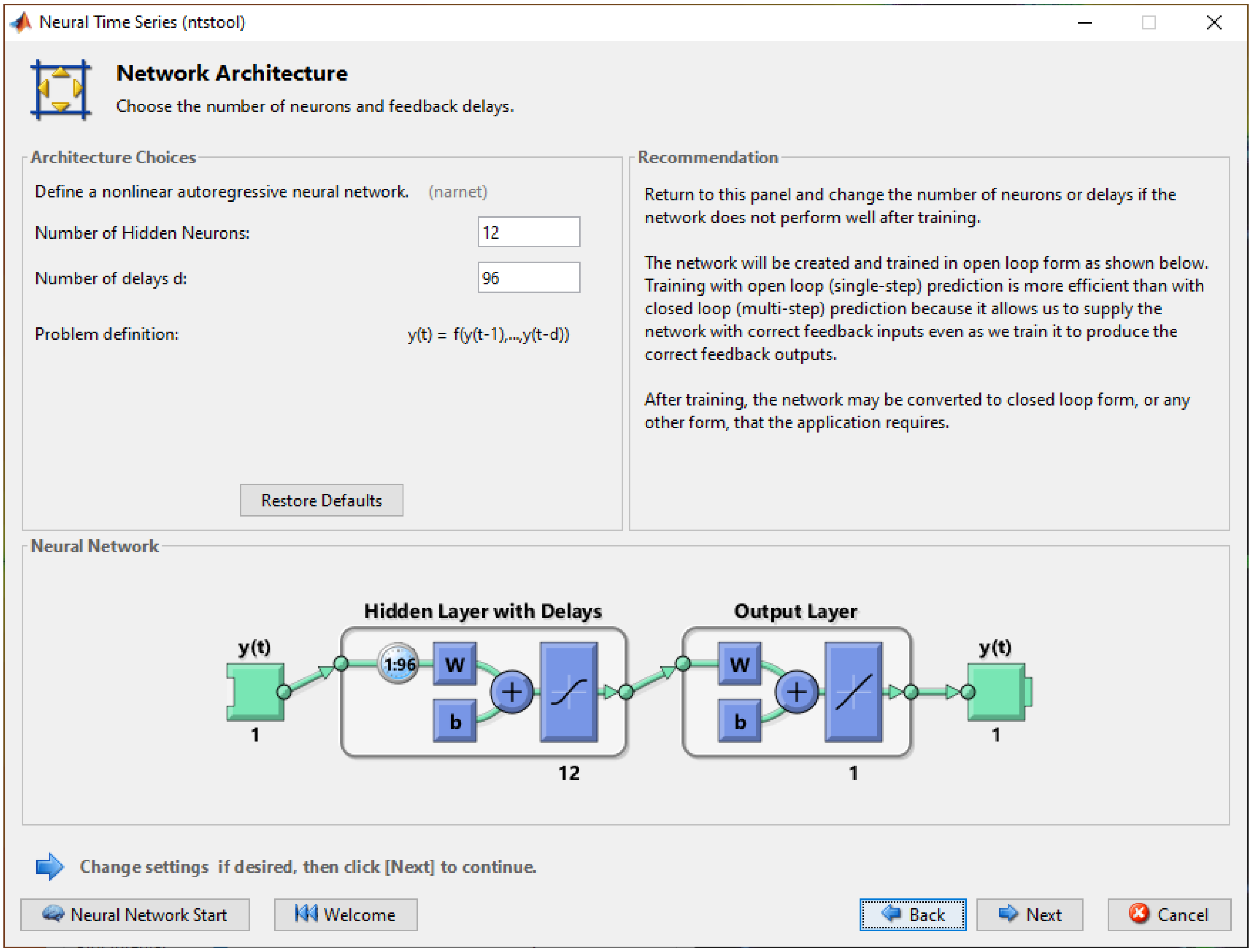

Figure 4. RNN NAR training in MATLAB 2015.

With the network trained, the following lines of code are executed using a MATLAB 'script', which allows propagating or making the forecast from the information trained by the RNN NAR.

```
T = tonndata(CAUDAL,false,false); % correct information for network
[x1,xio,aio,t] = preparets(net,{},{},T); % prepares the information for the type of network
[y1,xfo,afo] = net(x1,xio,aio); % spreads information over the network
[netc,xic,aic] = closeloop(net,xfo,afo); % generates a closed network from the previous one
[y2,xfc,afc] = netc(cell(0,96),xic,aic); % carry out the propagation 96 months or eight years
```

Data processing

Protocol for the implementation of ANN in precipitation models-runoff

Dawson and Wilby (2001) propose a protocol for the implementation of ANN in rain-runoff models, which consists of the following steps:

1° Collect data.

2° Select the prediction model

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

3rd Data preprocessing - stage 1: eliminate jumps and trends, if necessary, and remove seasonality. Select the variables to predict and the variables that will make the prediction, and choose the most influential.

4 ° Choose a type of ANN: type of network, training algorithm.

5th Data preprocessing - stage 2: scaling the data according to the output range of the chosen trigger function. For this step, Equation **¡Error! No se encuentra el origen de la referencia.**):

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

Where:

Z_T = climbing series

Mz, mz = maximum and minimum value of series Y, respectively.

Ls, Li = upper and lower limits to adopt, respectively.

Y = value to be scaled.

6° Train the ANN.

7° Validate the ANN.

Tecnología y Ciencias del Agua

The pre-processed information in the first stage has been scaled. The parameters required to scale each of the variables towards the working range of the hyperbolic tangent function (-1 to 1) are shown in Table 1. The entire range of the function has not been used by the recommendation of the protocol as mentioned above, but the values have been scaled in such a way that there is a maximum of 0.9 and a minimum of -0.9 in each variable.

Table 1. Parameters to scale the variables to the working range of the hyperbolic tangent activation function.

Variable	Station	Mz	mz	Ls	Li
Precipitation	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. porcon	568.90	0.00	0.90	-0.90
	Huamachuco	333.80	0.00	0.90	-0.90
	Jesus	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
Temperature	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
Coverage	NDVI	0.52	0.31	0.90	-0.90
Flow	Puente Crisnejas	205.60	0.62	0.90	-0.90

Collection and processing of meteorological and hydrometric information

The meteorological stations are unevenly distributed within the basin and its surroundings. Those better spatially distributed in latitude, longitude and elevation, and that also have reliable records over long periods have been selected. The information has been compiled from the stations shown in Table 2. The stations are located as shown in Figure 6.

Table 2. Hydrometeorological stations.

Station	Coordinates		Altitude	Registration Period	Years of Registration	Variables		
	Longitude	Latitude	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 and 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 and 2014-2017	13			x

To execute the first stage of data preprocessing proposed by Dawson and Wilby (2001), outliers were filtered using Tukey's (Tukey, 1977) box plots, and jumps in the mean were verified and corrected using non-parametric statistics tools, as the test of accumulated deviations of Buishand (1982), as well as in the variance through the test of Fligner-Killeen (Fligner & Killeen, 1976). Also, trends were analyzed with the Mann-Kendall test Kendall (1975). All the previous process was carried out in R 3.4.0 language, with the Trend and Climtrends packages. The data filling was carried out with the HEC-4 model of the US Army Corps of Engineers (1971), which is based on multiple regressions between each month of registration and between stations. This first stage has been carried out following the flow chart shown in Figure 5.

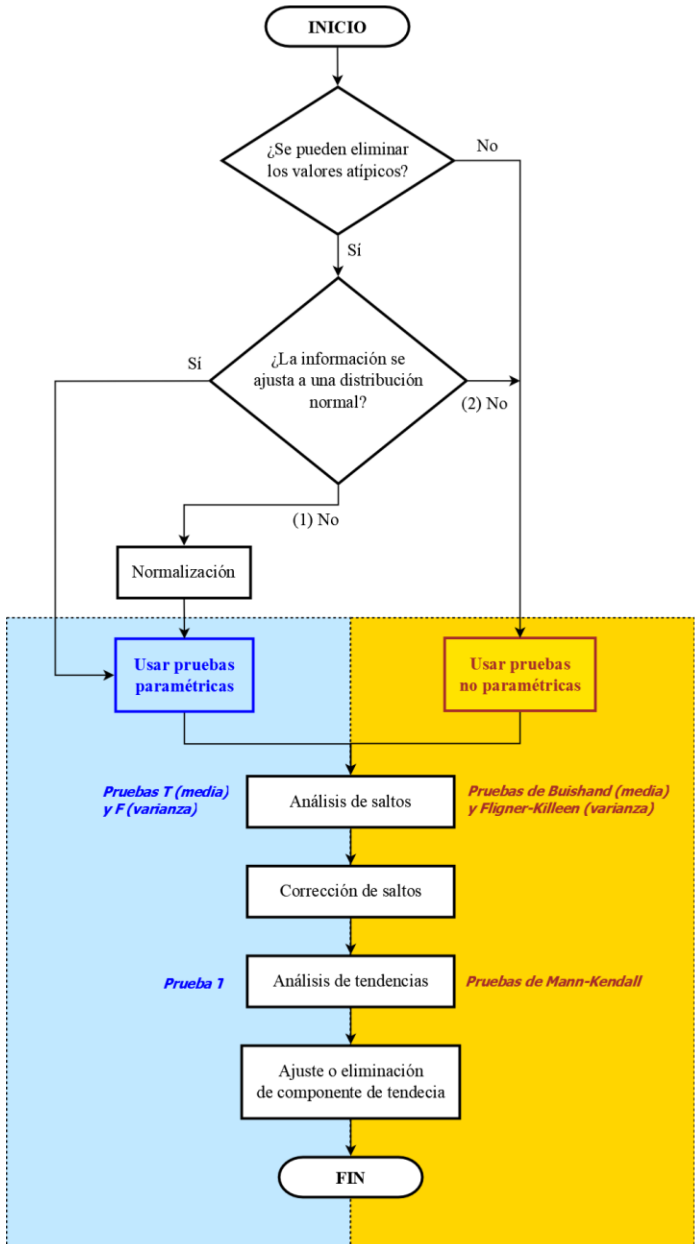

Figure 5. Hydrometeorological information pre-processing flow diagram.

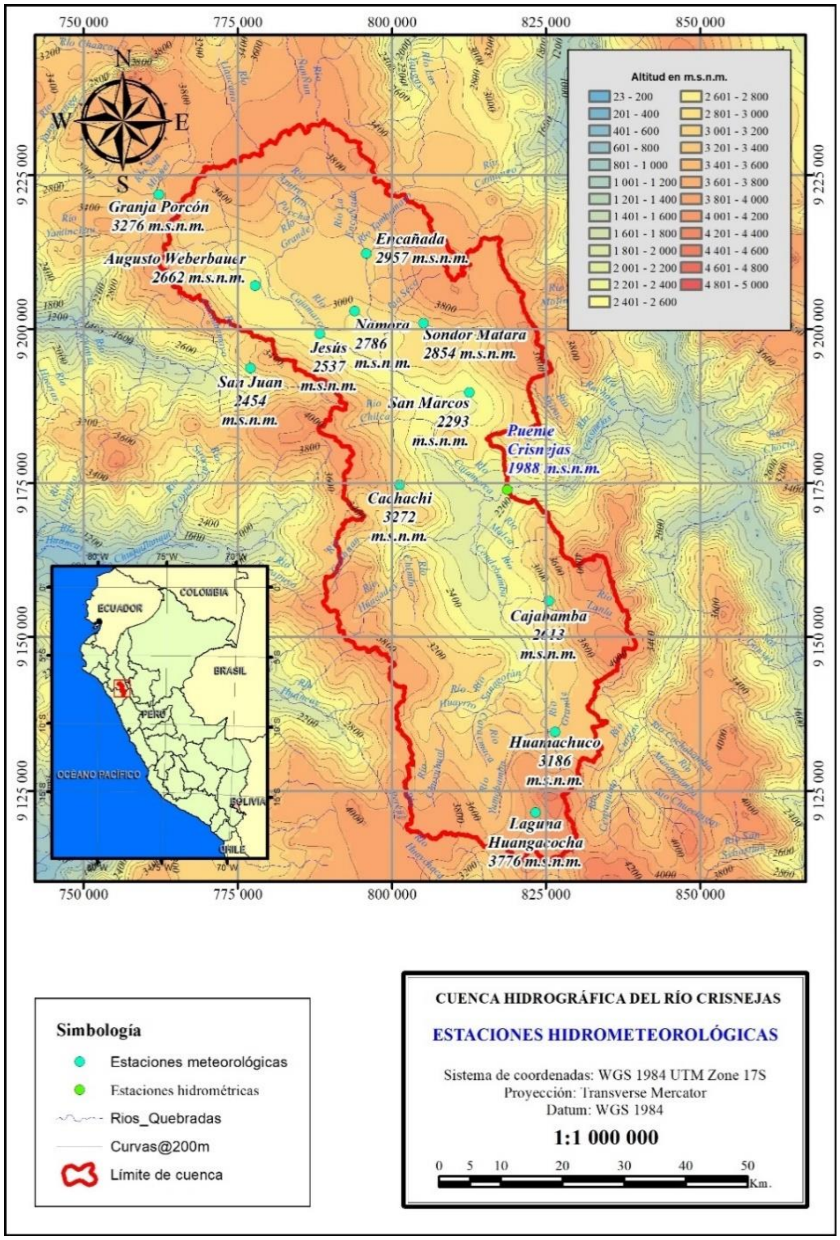

Figure 6. Hydrometric and meteorological stations.

Collection and processing of cartographic information

The basin has been delimited using an ASTER-GDEM digital elevation model. In addition, 12 multispectral images were acquired from the Landsat program corresponding to each month of the hydrological year, as shown in Table 3. These images have been used to determine the NDVI using the Tucker equation (Huete & Tucker, 1991):

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

Where:

NIR = band corresponding to the near-infrared.

Red = red spectrum band.

Table 3. Landsat images were used to determine NDVI.

Image	Year of taking	Satellite	Cloud cover

January	1992	Landsat 4-5	Less than 10 %
February	1992	Landsat 4-5	Less than 10 %
March	2016	Landsat 8	Less than 30 %
April	1995	Landsat 4-5	Less than 10 %
May	1995	Landsat 4-5	Less than 10 %
June	2011	Landsat 4-5	Less than 10 %
July	2005	Landsat 4-5	Less than 10 %
August	2007	Landsat 4-5	Less than 10 %
September	1984	Landsat 4-5	Less than 10 %
October	1986	Landsat 4-5	Less than 10 %
November	1998	Landsat 4-5	Less than 10 %
December	1991	Landsat 4-5	Less than 10 %

Prior to the calculation of the NDVI, the corrections and transformation of digital levels to physical parameters of each image were carried out, following the flow chart of Figure 7, adapted from Chuvieco (1996). The processing was done in QGIS 2.18, using the Semi-Automatic Classification Plugin.

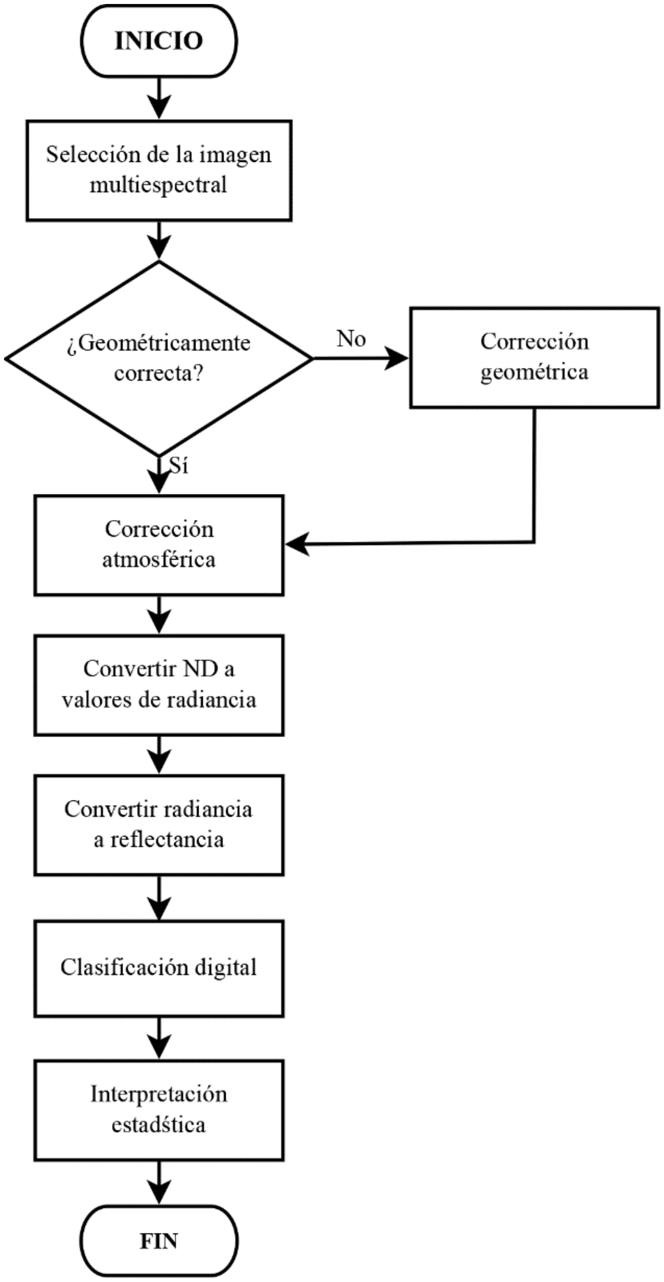

Figure 7. Multispectral Imaging Flowchart.

Study area

The Crisnejas river basin (see Figure 8) is located in northern Peru, in the departments of Cajamarca and La Libertad. The delimitation has been made from the point located on the Crisnejas bridge (Table 4), where a hydrometric station is installed that has recorded the river levels for more than 30 years but whose height-flow curves are not found available to transform this information into flows. There are only 13 years of daily flow measurements.

Table 4. Location of the Puente Crisnejas hydrometric station.

Point	Location				
	UTM-WGS 1984 Zone 17S		GCS WGS 1984		Elevation
	Eats	Nort	Latitude	Longitude	
Puente Crisnejas	818705	9173905	7° 27' 48.73"	78° 6' 47.25"	1988

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

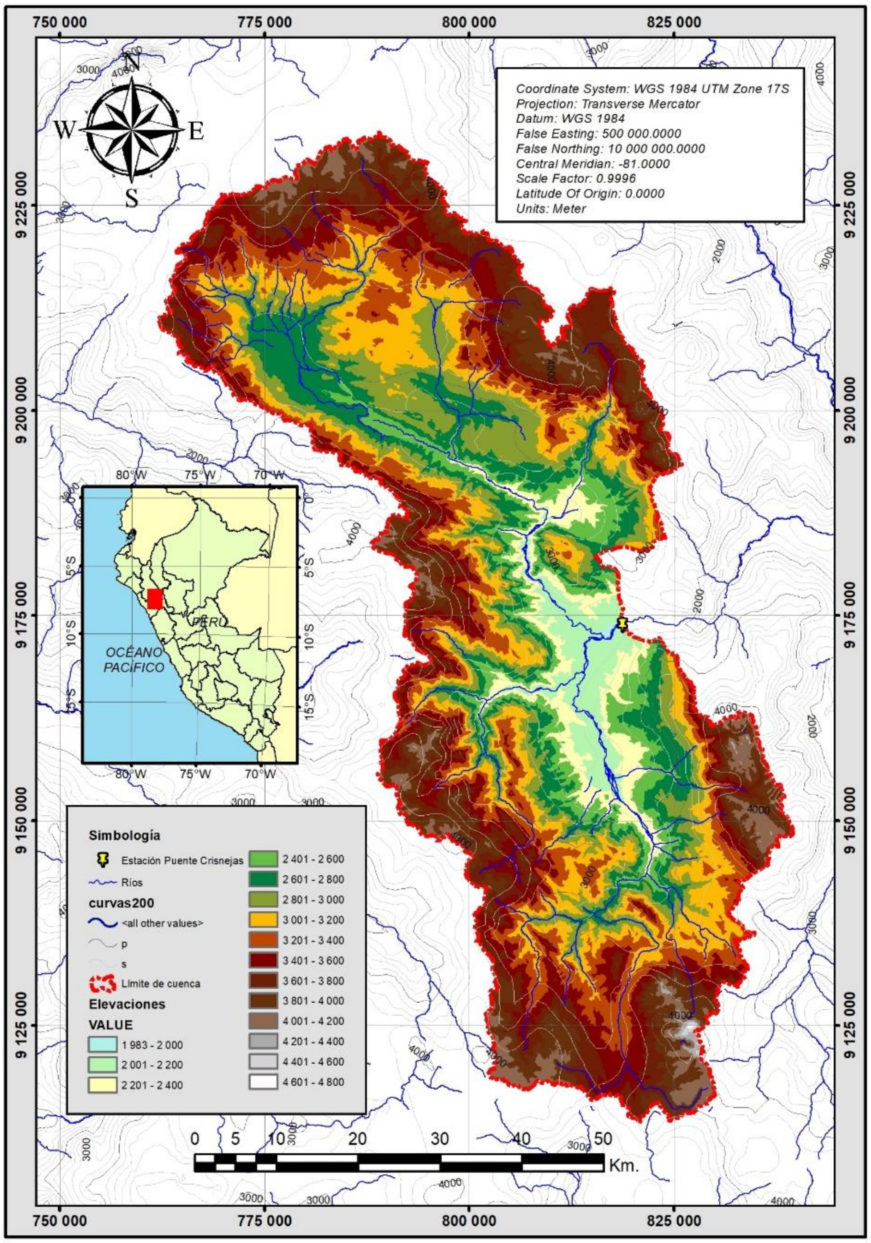

Figure 8. Crisnejas river basin.

Results

Analysis of cartographic information

In addition to delimiting the basin, the normalized difference vegetation indices (NVDI) have been determined for each month of a hydrological year assumed as an average, as shown in Table 5 and Figure 9. In some months, cloud cover did not allow obtaining the NDVI in some areas of the basin, however, since the required numerical data is an average, Figure 10 shows the spatial distribution of the NDVI. information was not completed, and only the average of what was captured in the survey was obtained from the image.

Table 5. NDVI, the monthly average for the training of ANN MLP.

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
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

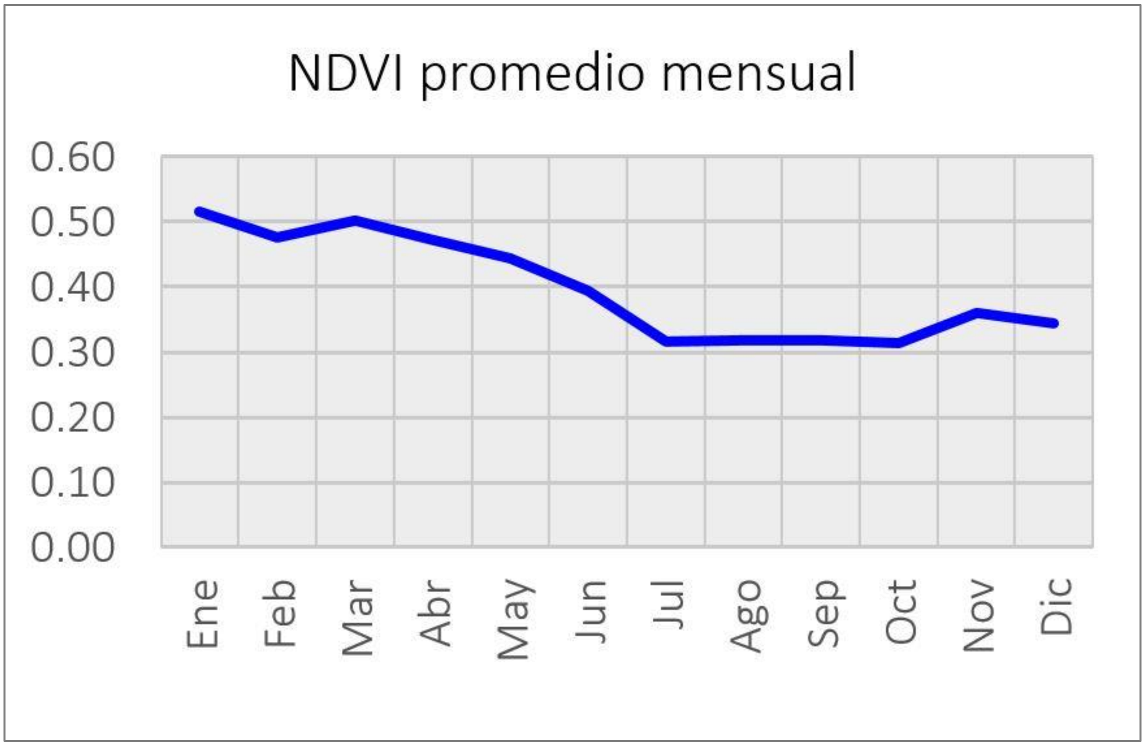

Figure 9. Normalized Difference Vegetation Index - NDVI, monthly average.

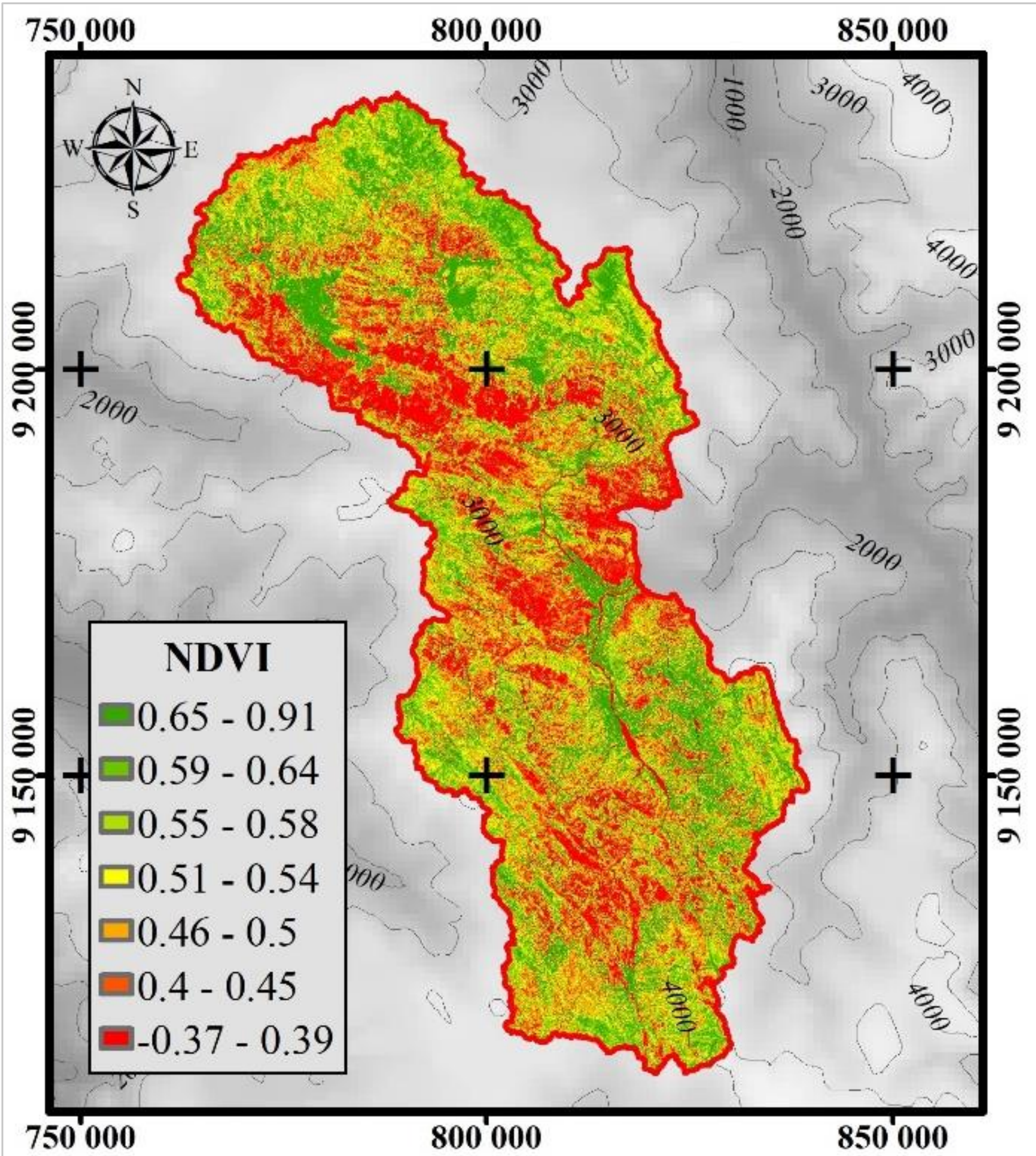

Figure 10. NDVI calculation.

Analysis of hydrometeorological information

The processing of the hydrometeorological information resulted in obtaining time series of precipitation and monthly temperature homogeneous both in the mean and in the variance and free of trends and atypical values. In addition, the record of all meteorological stations was standardized by extending the short record time series (Figure 13 and Figure 14).

In general, the behavior of the hydrological cycle in the region shows a wet season from September to March and a dry season from April to August, as shown in

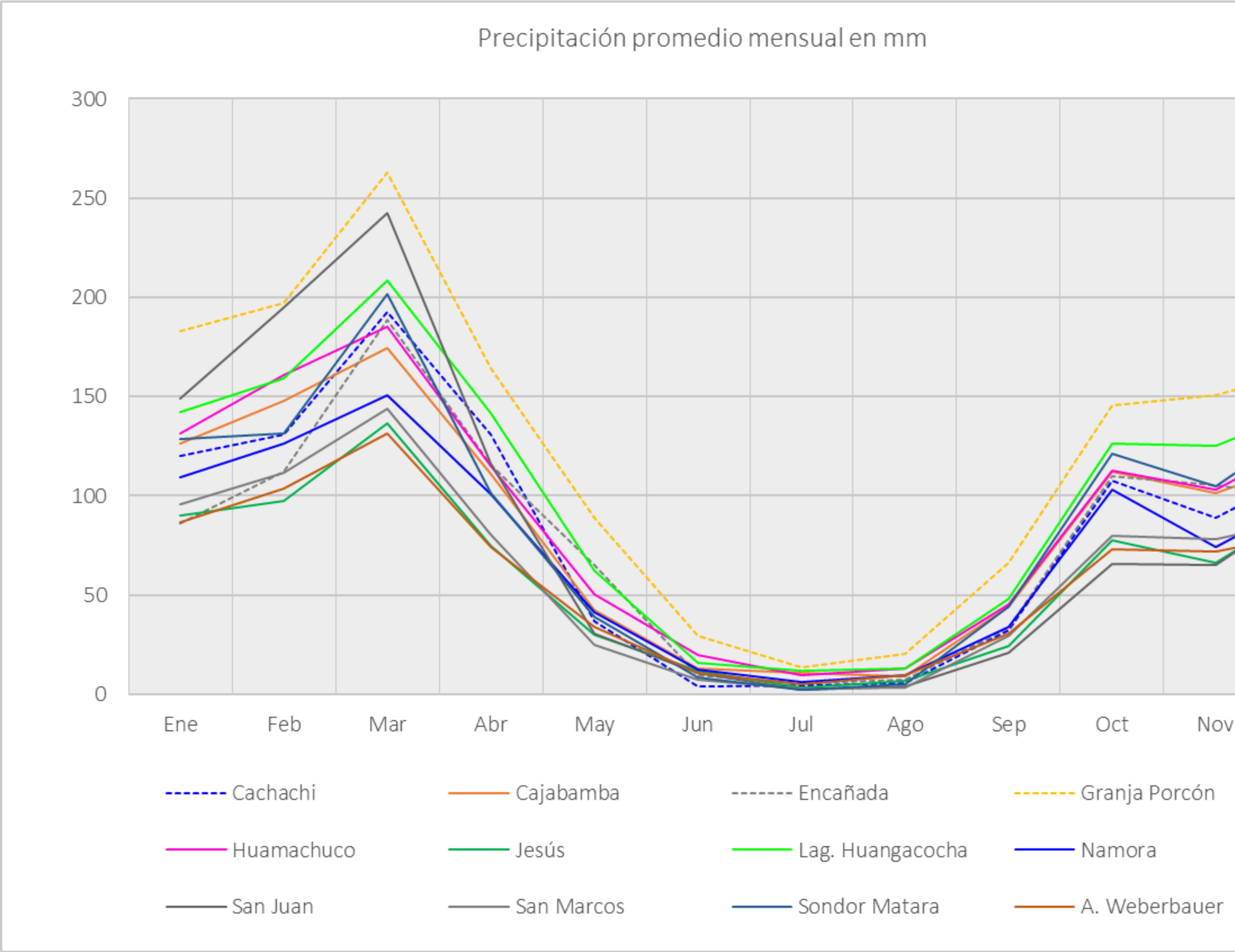

Figure ***11***. The temperatures show behavior with higher values. High in the wet season and lower in the dry season, except for the San Juan station, where the reverse occurs (see Figure 12). Initially, it was thought to discard this station. However, it was not eliminated since its behavior

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

could enrich the behavior of the ANN MLP; if not, it is the weights in training that rule out its influence.

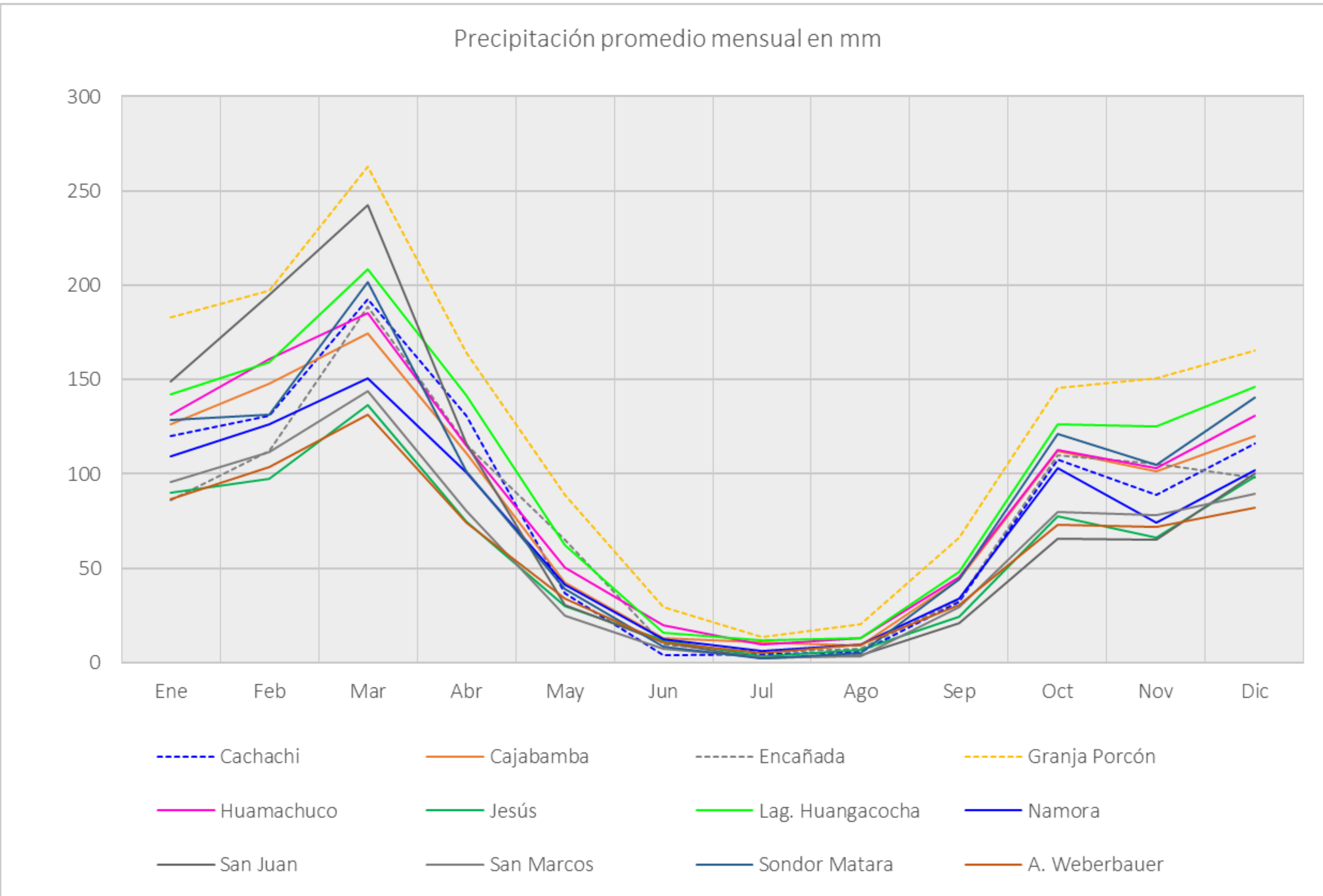

Figure 11. Average monthly precipitation in mm, for complete and extended records in 1965-2017.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

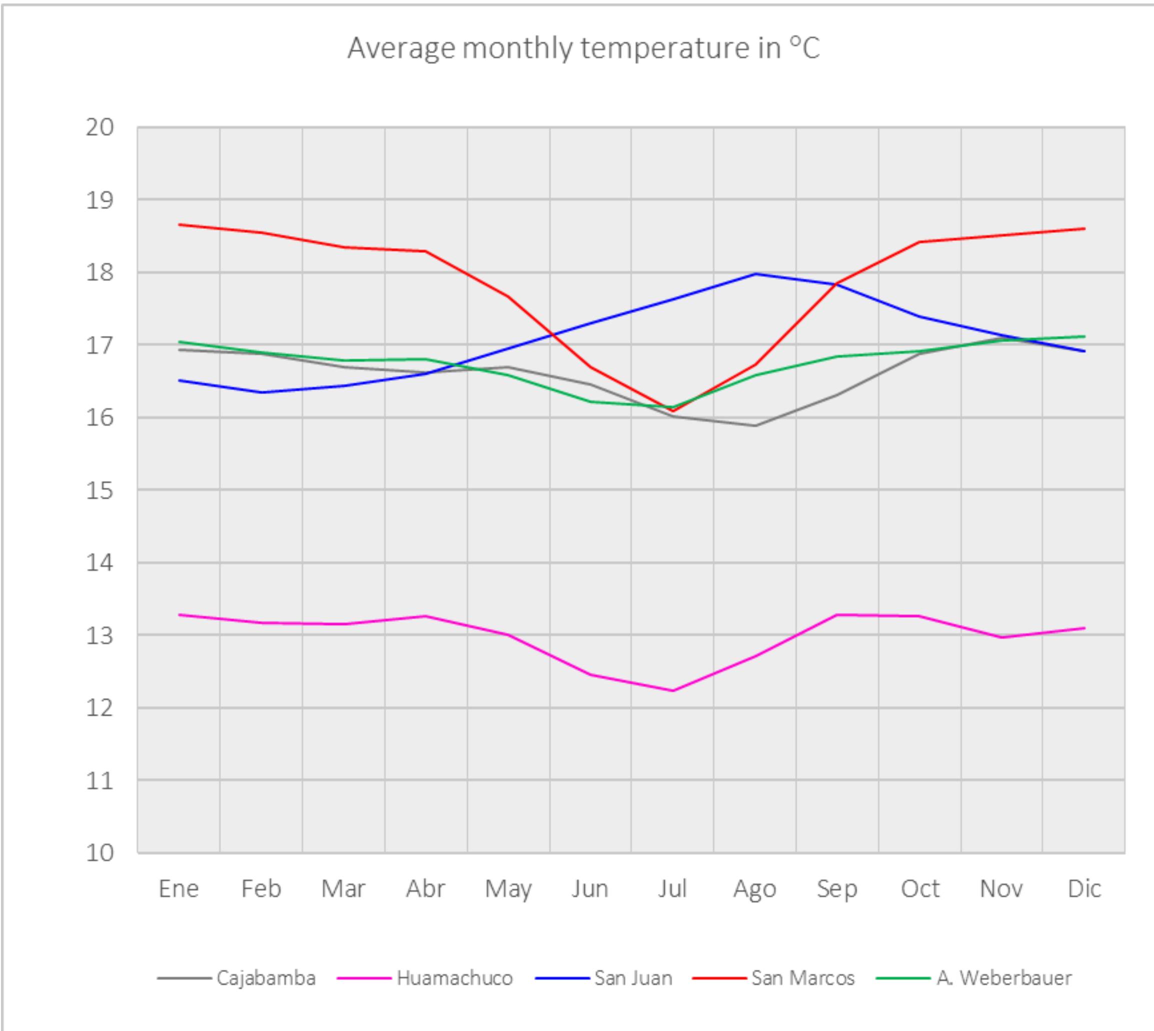

Figure 12. The average monthly temperature in °C, for records from 1965-2017.

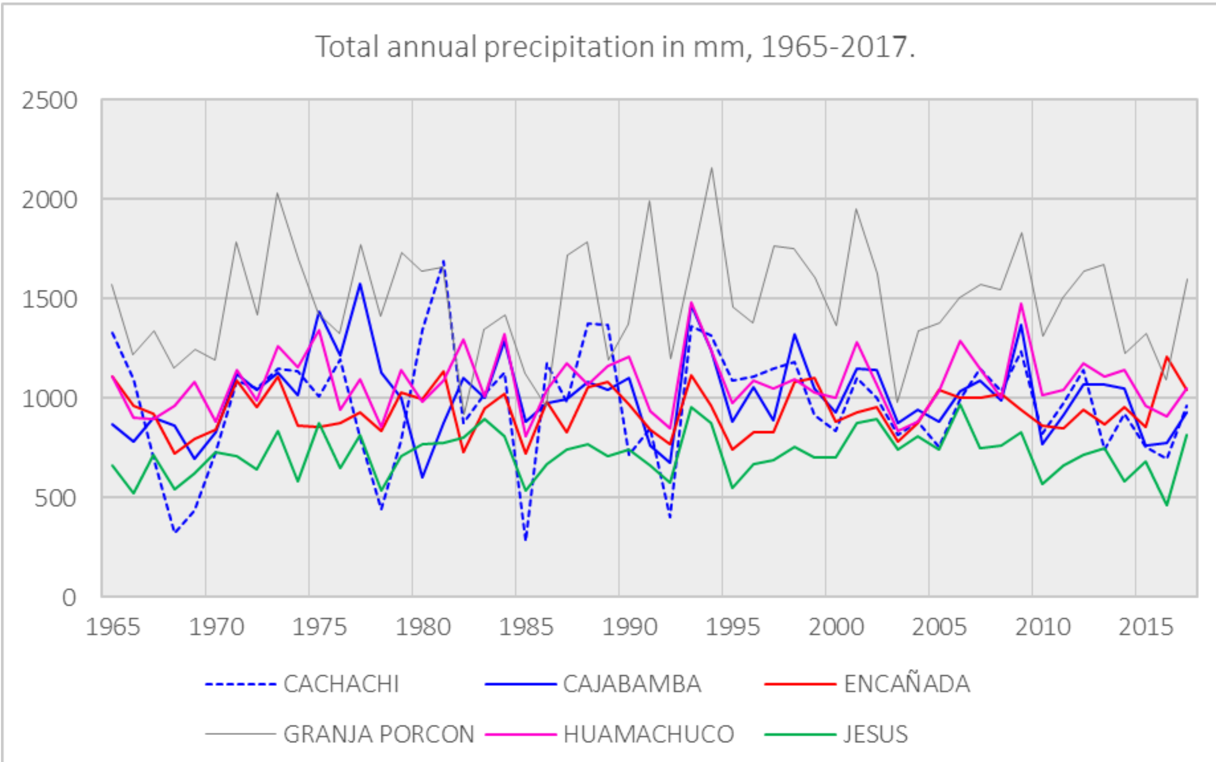

Figure 13. Total annual precipitation in mm, 1965-2017.

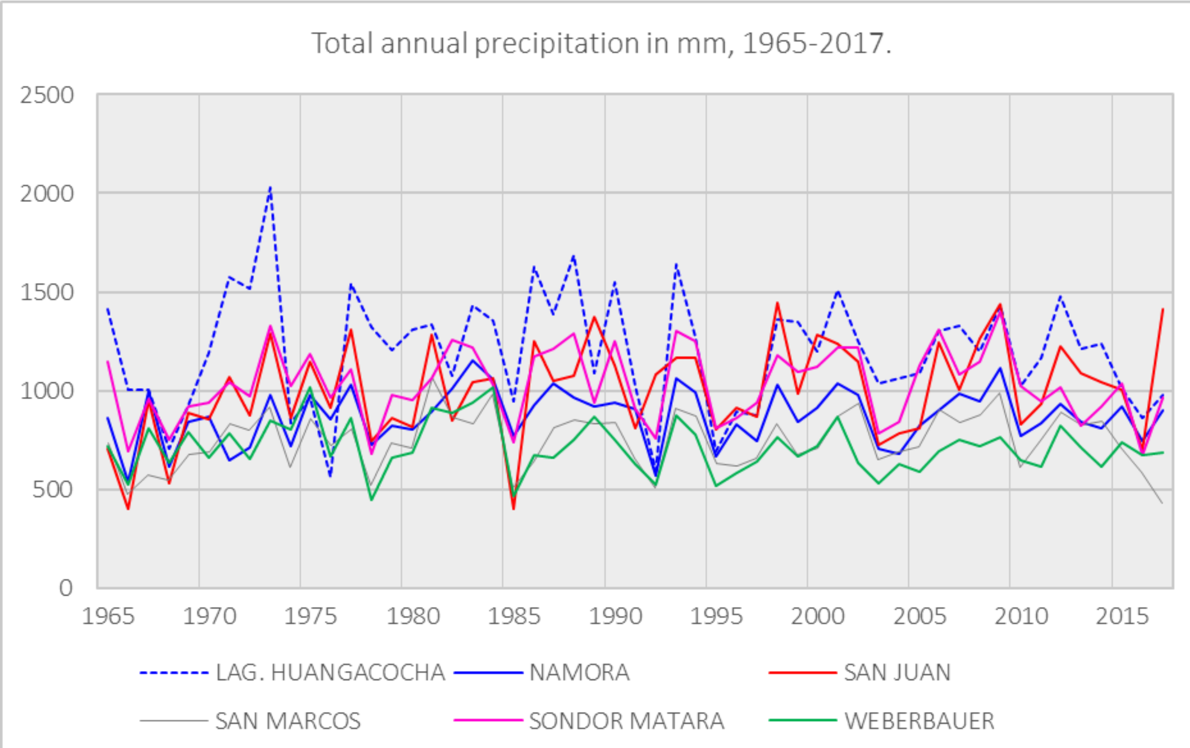

Figure 14. Total annual precipitation in mm, 1965-2017.

The application of non-parametric statistics tools has allowed the time series analysis to be more reliable and consistent with the expected hydrological behavior of the studied region.

The same procedure has been followed for the flow analysis of the Puente Crisnejas hydrometric station (Figure 15).

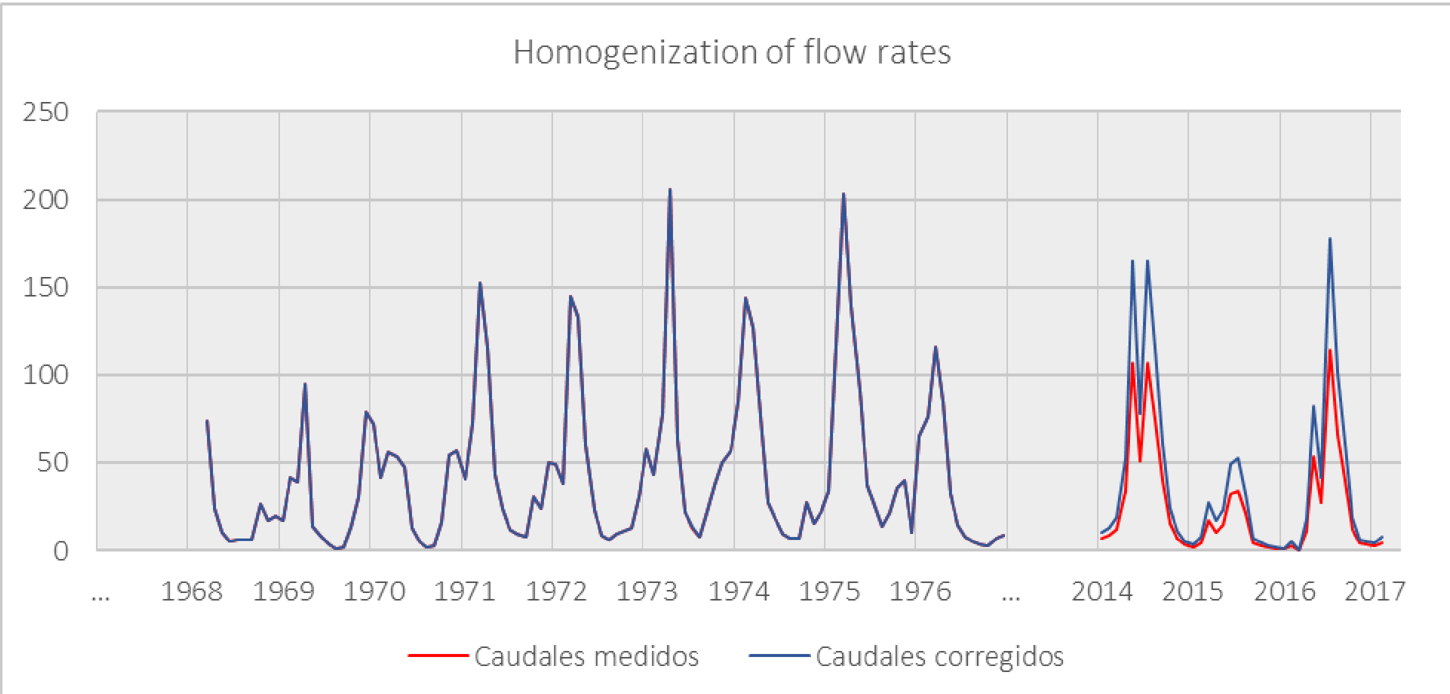

Figure 15. Homogenization of flows (m3 / s) at Puente Crisnejas station.

Training of the multilayer perceptron (ANN MLP). Estimation of flows in historical record 1965 to 2017

The ANN MLP training shows a high fit between the measured data and the data trained by the network, as seen in **¡Error! No se encuentra el**

origen de la referencia.. Validation has been carried out between the reserved data of the training to evaluate the predictive capacity of the network for untrained employers. As expected, the data from trained patterns (**¡Error! No se encuentra el origen de la referencia.**) present, in general, a better fit with the measured than the information generated from untrained patterns (**¡Error! No se encuentra el origen de la referencia.**). Even so, said information shows a high degree of goodness of fit according to the measures or coefficients considered by Cabrera (2012).

Table 6. The goodness of fit of flow rates estimated by MLP-type ANN

Goodness-of-fit measures[1]	**Trained period (1968-1976, 2016)**		**Untrained period (2014, 2015, 2017)**	
	Value	**Qualification**	**Value**	**Qualification**
Calibration coefficient (**r**)	0.99	Correlation strong positive	0.90	Correlation strong positive
The determination coefficient (**r^2**)	0.97		0.81	
Schultz coefficient(**D**)	1.38	Very good	8.65	Good
Cumulative mean deviation (**MAD**)	6.25	-	18.92	-

Nash-Sutcliffe efficiency (**E**)	0.96	Excellent	0.77	Very good
Mass balance error (**m**) in %	11.29	-	-3.08	-
Root mean square error (**RMSE**)	8.69	-	28.55	-

[1]Cabrera (2012).

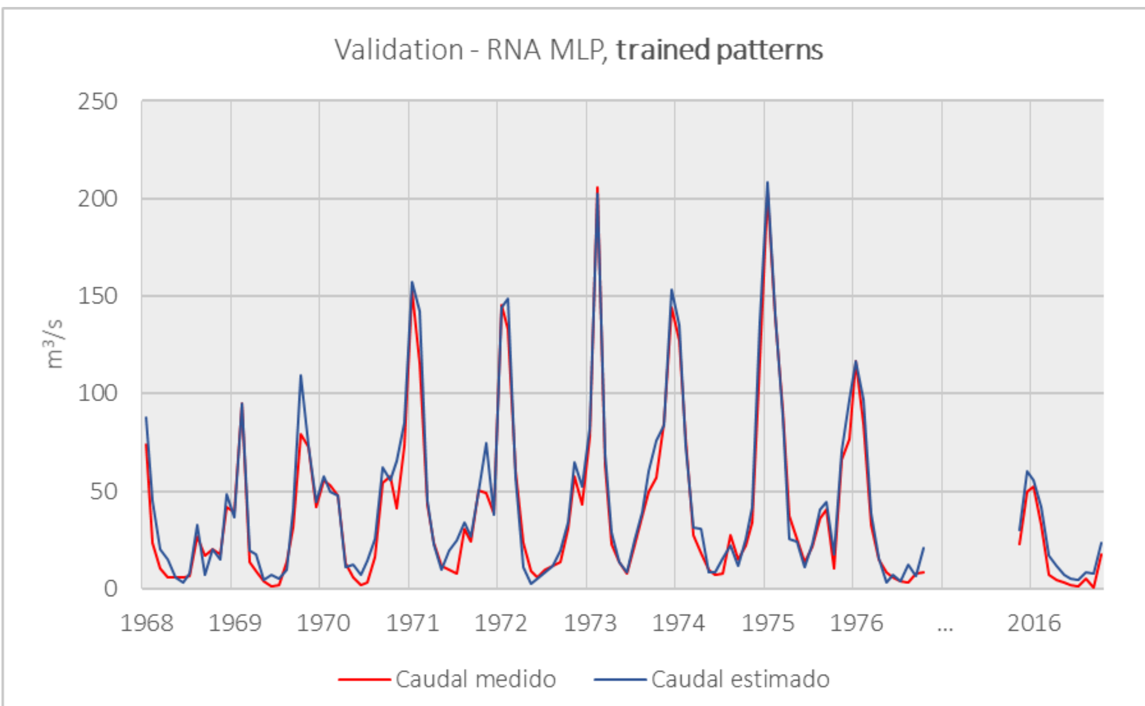

Figure 16. Learning - ANN MLP for trained monthly patterns.

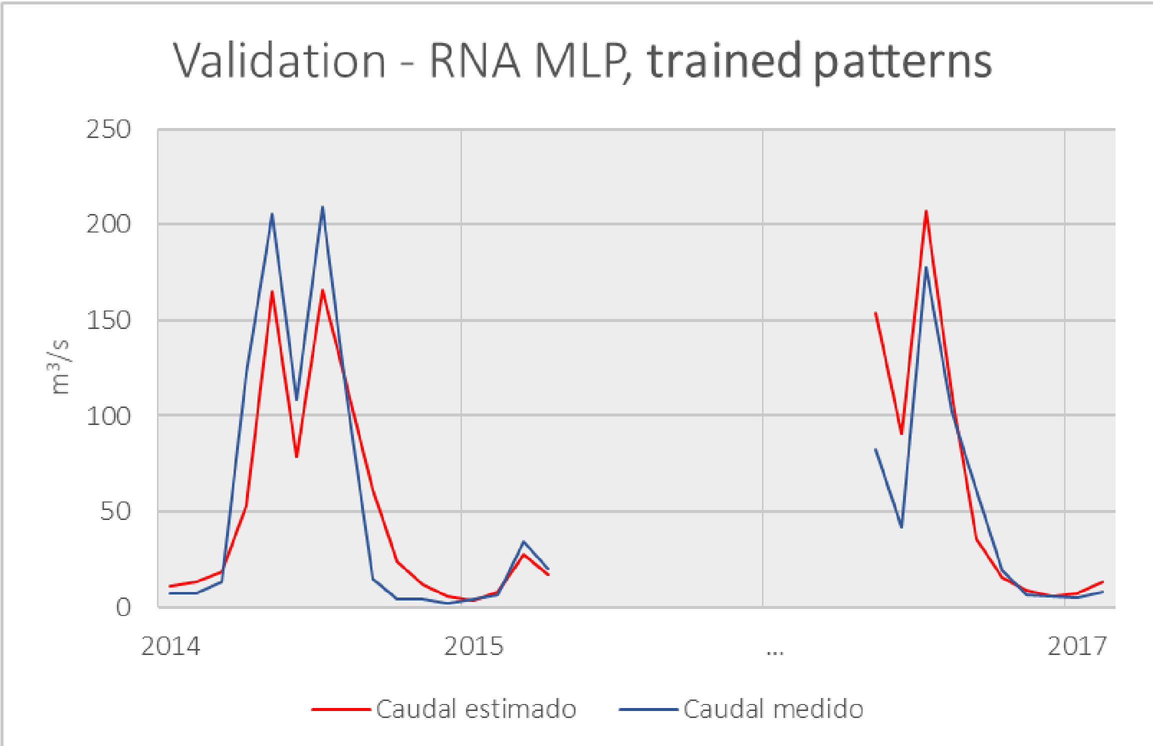

Figure 17. Validation - MLP ANN for untrained monthly standards.

As seen in the slopes of the regression lines, **¡Error! No se encuentra el origen de la referencia.** indicates a good fit between the information measured and that estimated with the MLP ANN, both for trained and untrained patterns.

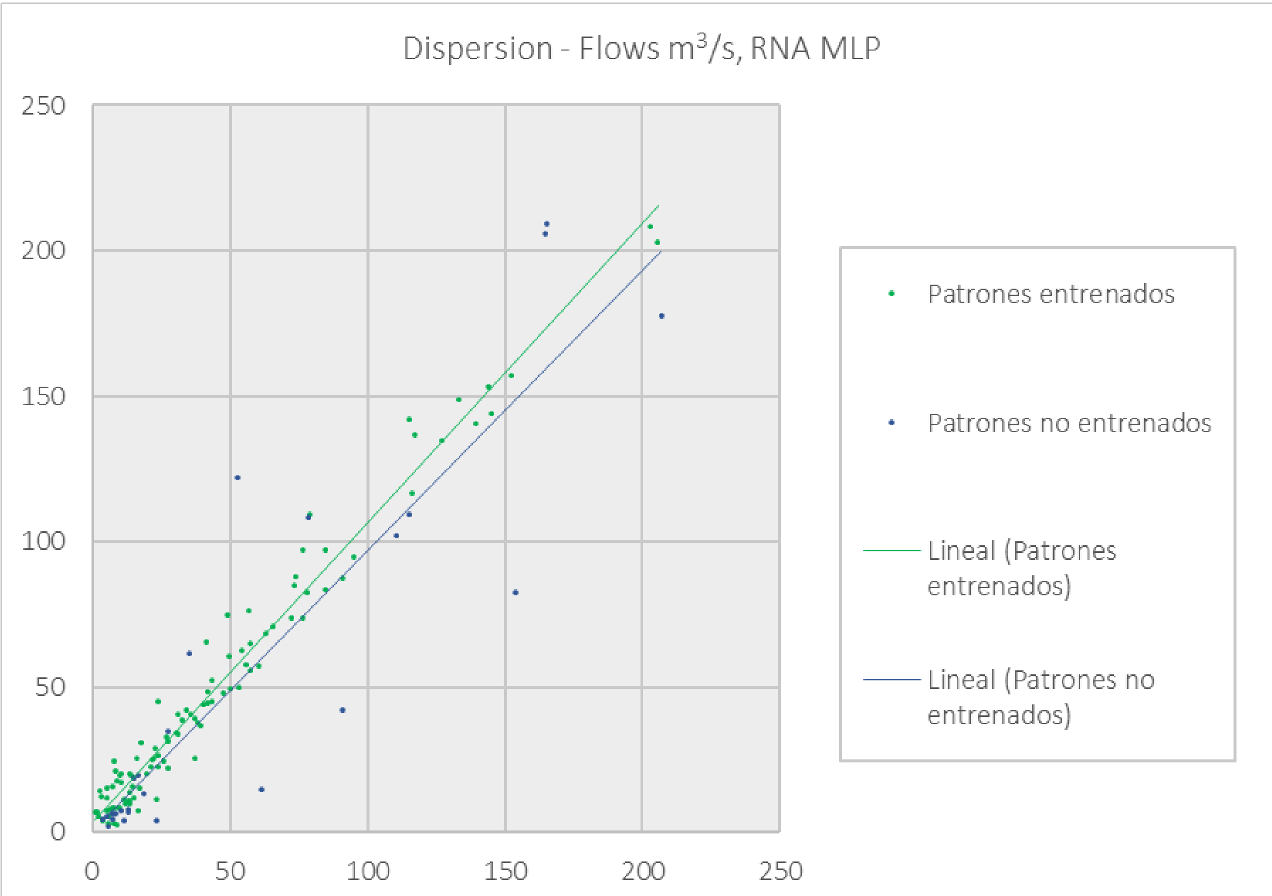

Figure 18. Dispersion of monthly flows, ANN MLP.

Regarding the assessment of the accumulated mean deviation (MAD), it is important to clarify that this parameter is intended to be as close to 0 as possible since it represents the average of the differences between the observed and estimated data. The value of 6.35 of the data of the trained period and 18.92 of the untrained period can be directly interpreted as the "average error" in m^3/s between the estimated information and the measurement in said periods.

The mass balance error (m) represents, in quantity, the relationship between the volume of the observed hydrograph and the simulated one. In the same way, it has a better evaluation the closer it is to 0. In this case, there is less error in the data generated for the untrained period.

The root means square error (RMSE) quantifies the magnitude of the deviation between the measured and estimated values; similarly, a value closer to 0 implies a better fit. Again, the trained period presents a better fit for this particular case than the untrained period.

The record of monthly flows generated with the MLP ANN for the period between 1965-2017 is shown in Figure 19.

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

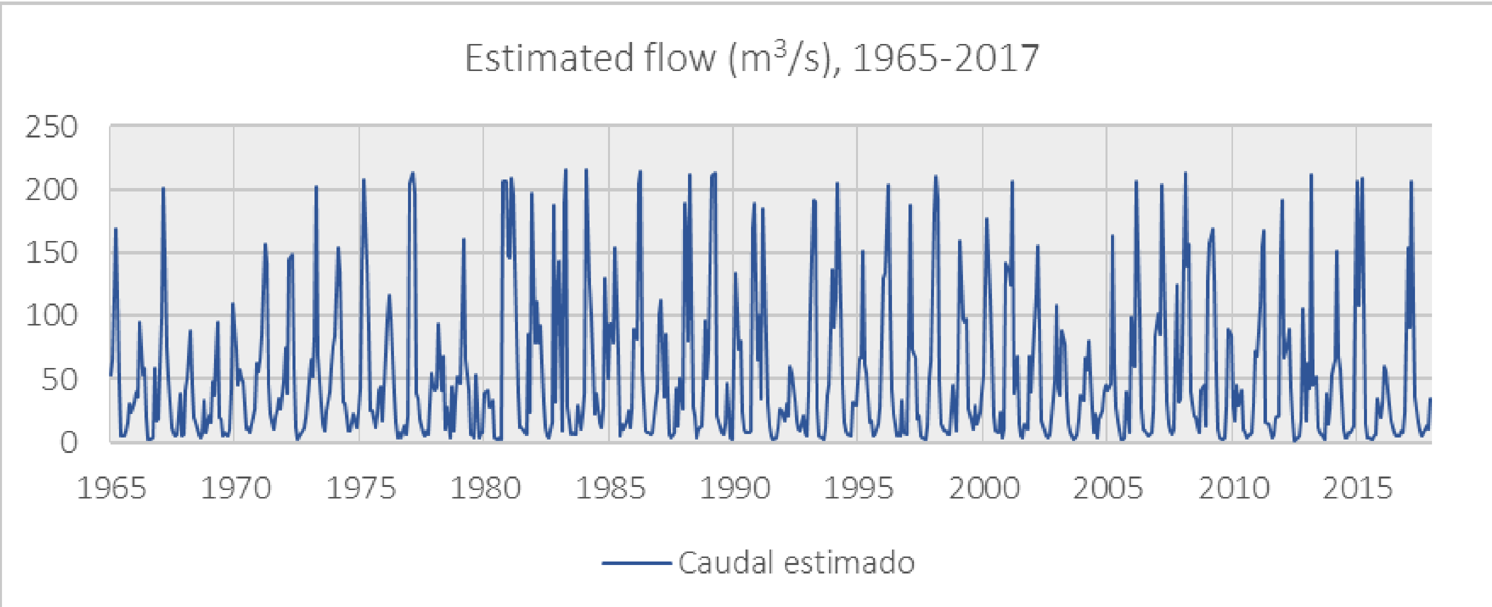

Figure 19. Record of monthly flows (m^3/s) estimated with the ANN MLP, 1965-2017.

Recurring network training (RNN NAR). Flow forecast in the record from 2018 to 2025

The monthly flow data estimated with the MLP ANN were used (in its scaled form) to train the RNN NAR, which outputs the projected monthly flow data until 2025, as shown in Figure 23.

The forecast analyzed jumps and trends before being taken as valid, having to discard the response given by several trained networks. Finally, the data that did not need to go through corrections of this type were selected.

It should be mentioned that, given the nature of the network and the noise in the data, some negative values are usually generated that are absurd in the forecast. In this case, there were six values of the 96 months. However, they were purged and replaced by the immediately higher positive value. To validate the forecast, the network had to be trained many times until it learned almost perfectly the behavior of the monthly flows' time series to reduce errors in the forecast. The training results generated in MATLAB are shown in Figure 20 and Figure 21. The correlation value in the training period is almost perfect. The validation done by MATLAB also indicates a moderate positive correlation. Monthly data from the last two years were reserved to determine the precision of the forecast; Table 7 shows the goodness of fit.

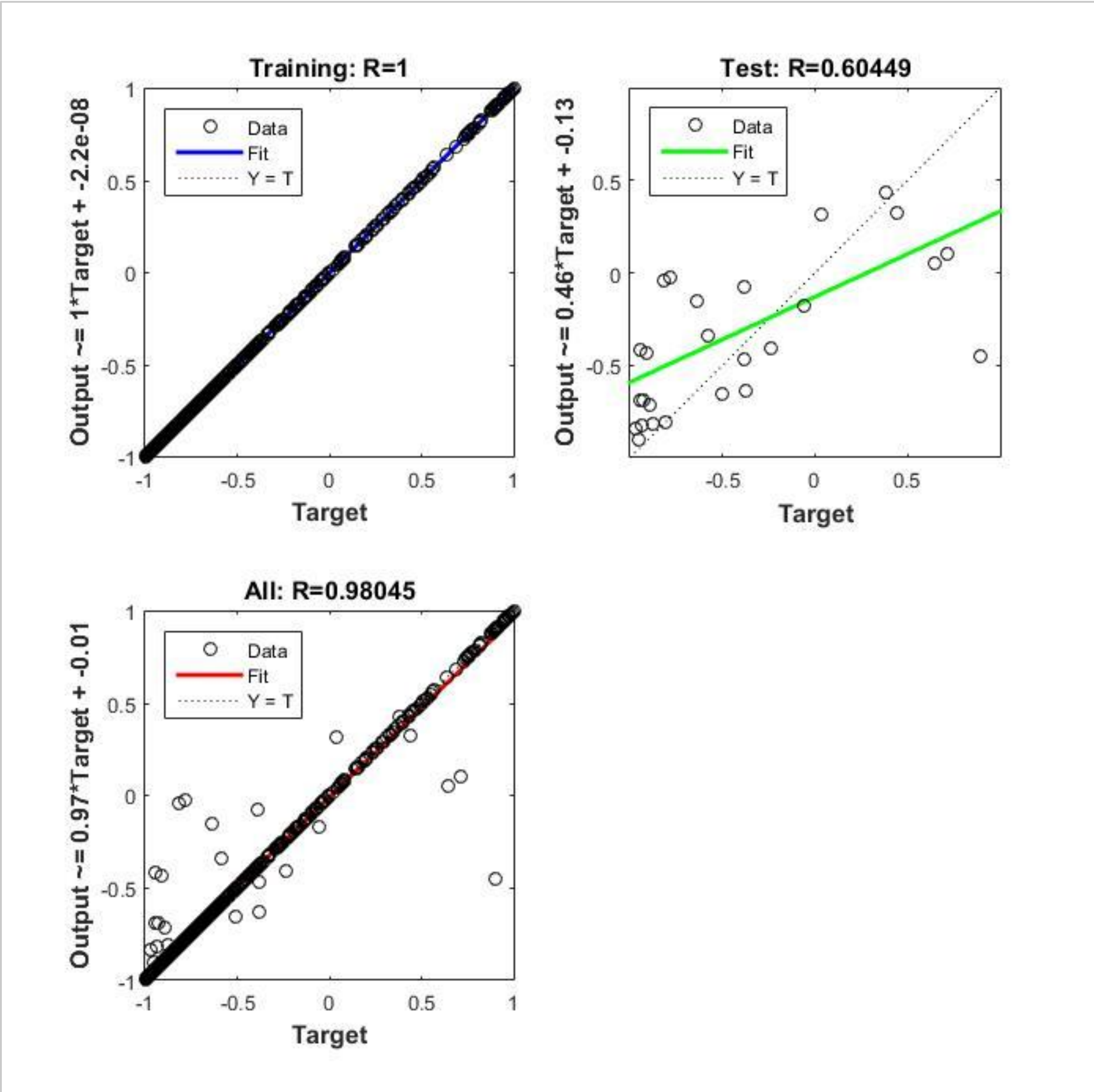

Figure 20. Correlation coefficients determined by MATLAB.

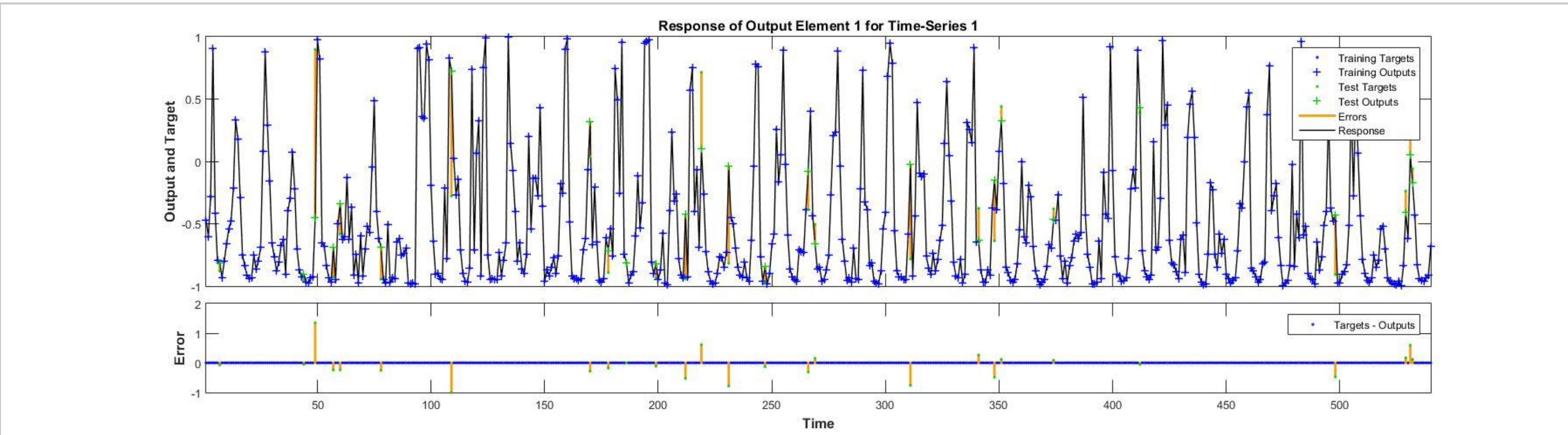

Figure 21. RNN NAR response for the trained time series.

Table 7. The goodness of adjustment of flows estimated by the ANN type NAR, comparative of periods measured and forecast 2017-2019.

Goodness-of-fit measures	**Flow forecast (2018-2025)**	
	Value	**Fit**
Calibration coefficient (**r**)	0.84	Positive - Strong
The determination coefficient (**r^2**)	0.71	
Schultz (**D**)	8.25	Good
Cumulative mean deviation (**MAD**)	13.61	
Nash-Sutcliffe efficiency (**E**)	0.64	Very good

Mass balance error (**m**)	33.55	
Root mean square error (**RMSE**)	23.55	

In general, the data present an acceptable fit, considering that they are forecasts and their value can always be affected by variables not controlled in the simulation (demand growth or climate change) and training of the RNN NAR.

The forecast could be less certain the further it is from the last measured data, given that the forecast error becomes larger with each step of the propagation, taking into account that each data generated depends on the last 96 data, which supports the reason why the forecast of a long period with this type of technique is not convenient.

Forecast data is displayed in Figure 22 and the complete series is shown in Figure 23.

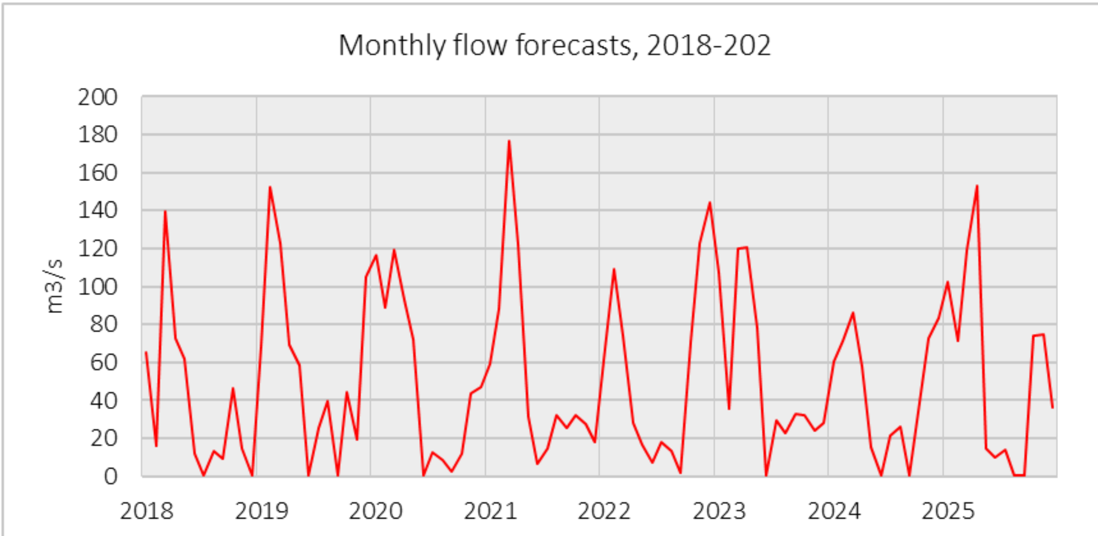

Figure 22. Forecast of monthly flows, period 2018-2025.

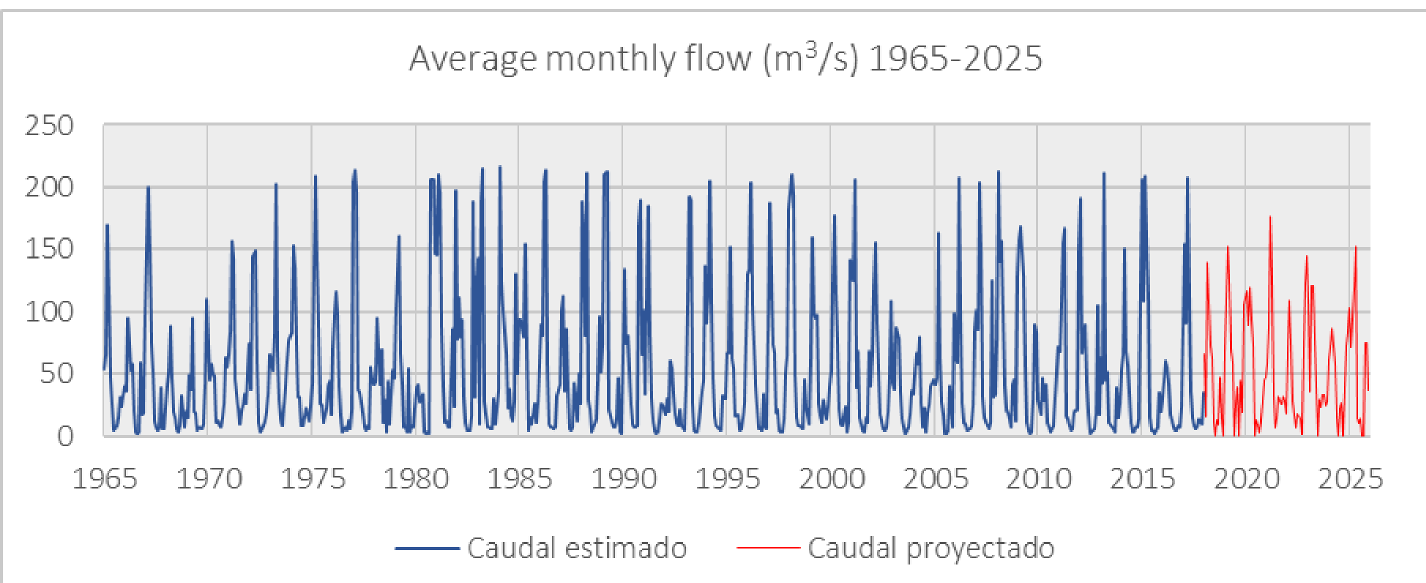

Figure 23. Monthly flow forecast at Puente Crisnejas station, extended until the year 2025.

OPEN ACCESS

The annual average of flows is shown in Figure 24 to observe a summary of the behavior predicted by the RNN NAR.

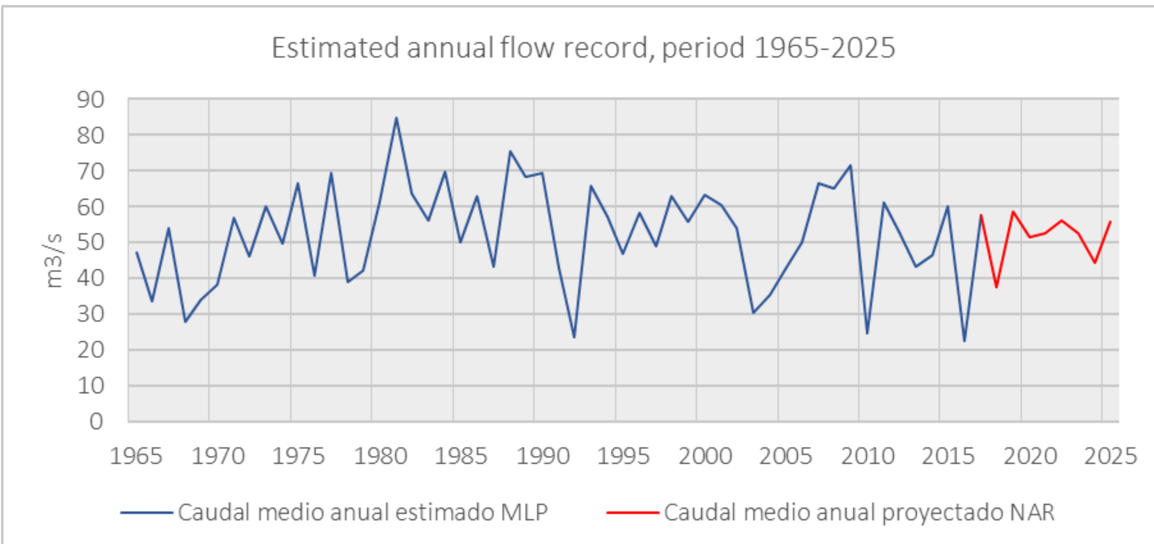

Figure 24. Estimated record of the mean annual flow, period 1965-2025.

OPEN ACCESS

Discussion

The artificial neural networks used in the research have allowed the estimation of the missing monthly flow record and the forecast of these flows, resulting in a total synthetic series of 61 years of record (1965-2025). This registry provides a better overview of the river's water supply for planning and developing future water use projects.

The complete record does not show a significant trend in the data; however, the forecast series shows low flow values. This could be due to the errors in the measurement of the initial years with which the multilayer perceptron was trained; even when the information was corrected for jumps, the variation is noticeable between the information measured in 1968-1976 with the period 2014- 2019. Unfortunately, this factor cannot be controlled, given the lack of metadata in the hydrometric station.

Despite the above, the results of this research demonstrate the robustness of recurrent multilayer perceptron-type artificial neural networks (ANNs) in the generation of synthetic series of monthly flows from meteorological information with a high goodness of fit. In turn, an adaptable procedural basis is shown for its extrapolation in basins with a similar record of information and even for cases in which a better temporal resolution is required, such as daily, a result that is compatible with those found by Lama and Sánchez (2020), who evaluated the effect of decomposition techniques to use them with a recurrent neural network called short-term long memory to increase the precision of the daily prediction of the Chira river flow in northern Peru. Likewise, Lee, Lee, and Yoon (2019) and Heras and Matovelle (2021) obtained prediction results that showed good performance with minimum mean square errors with high correlation coefficients, ensuring that the ANN models are suitable for evaluating complex hydrological and hydrogeological water systems.

The technique used has made it possible to use the largest amount of measured and available information on the basin without having to resort to preliminary simplifications in the variables (estimation of other variables using empirical equations) of the hydrological cycle and resulting in a complete record with a high goodness of fit.

Using non-parametric statistics tools has made it possible to simplify information analysis. It has not been necessary to resort to normalizations or other techniques that give validity to the data to be applied with traditional statistical tests. It is important to bear in mind that you have worked with a relatively large amount of data and that, in future research or work that requires a better temporal resolution, the amount of information to pre-process before training could be very complex if it is not considered this aspect.

Other research works on ANN for the generation of synthetic series of monthly flows, such as that of Laqui (2010), shows that a scheme based on an MLP ANN with current and antecedent precipitation and evapotranspiration inputs shows a better correlation between what is measured and estimated than just with current precipitation and evapotranspiration data, for your case. This is not necessarily decisive in all basins, considering the delay of each one or other factors that could influence the monthly hydrological behavior. As no other conceptualization of the basin has been investigated in terms of its variables, the training has been improved, modifying the conFiguretion parameters of the network, such as the number of layers or neurons and even the activation function, and an even higher correlation coefficient has been obtained. Gomes-Villa-Trinidad (2016) applies the neural networks in the flow forecast of the following month using an MLP ANN. However, since its objective differs from this research's, the conceptualization of the training patterns is also different. In its case, it uses the flow of the previous historical month and the rainfall and temperatures, achieving good results in predicting the flow for the following month. However, as previously said, the generation of a synthetic series is not sought but rather a forecast. The forecast for the present research was carried out historically and with another type of network architecture (RNN NAR) since this synthetic record allows a useful long-term visualization in decision-making.

The MLP ANN scheme trained in this research is a good starting point for future research that requires the generation of synthetic series of monthly flows. It is important to point out at this point that, unlike other research, here it is not has carried out transformations between measured variables; the variables taken in the field have been those that trained the MLP ANN, which shows the advantage of artificial neural networks in terms of taking advantage of the greatest amount of information measured in the basin.

As seen in the results, neural networks and satellite information have wide applications in estimating records and forecasting flow in the short or long term. The researcher's understanding and adequate selection of variables in the study only limit them to the process that requires modeling. Well, Herrera *et al*. (2020), like us in their research, propose models based on artificial neural networks and satellite information for filling in missing data in meteorological stations and spatial reconstruction of precipitation and temperature variables for the region of the Department of Valle del Cauca, Colombia, with results obtained that reach correlation coefficients of around 0.9.

Future research could also analyze the trained weights, determine the influence of field measurements at each station with respect to flow, and even try to interpret the behavior through regional equations.

Conclusions

The generation of the historical and forecast series of flows through training of artificial neural networks has been satisfactory and with a high goodness of fit, which allows us to have a solid base in terms of decision-making in future projects of water use of the basin.

This work shows the technique's robustness and high capacity for adaptation and use of the information measured in the basin. A protocol adaptable to basins with similar hydrometeorological records has been shown, such as a large number of basins on the Peruvian coast and highlands that otherwise would have to resort to precipitation-runoff models that do not always give at least acceptable results or that require lengthy calibration processes or additional field measurements concerning the parameters required by each model.

In addition, a scheme and conFiguretion of ANN MLP and RNN NAR are presented as a starting point in similar analyses.

The methodology used can be extrapolated to many cases since techniques have been used for the analysis, correction, and processing of meteorological data that are characterized by their wide range of application in different types of data, in this case, non-parametric statistical techniques and artificial neural networks, for which there are multiple free-to-use tools. In addition, these techniques give good adjusted results without the need to resort to assumptions or make assumptions about the data, and they have not had to resort to calibration processes.

Finally, the information provided by this research shows the feasibility of using artificial neural networks to estimate synthetic series of monthly flows, both in historical records and in forecasts.

References

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*. Recovered from 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., Rezanezhad, F., Roshan, M. H., Malekian, A., & Shamshirband, 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). *Boletin 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 minirrobots*. 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 Amambaí (Brasil)*. Recovered from 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, Recovered from 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. Recovered from 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-aportación*. 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. Recovered from https://doi.org/10.26507/ponencia.273.

Tukey, J. W. (1977). *Exploratory data analysis*. Vol. 2 (pp. 131-160). Recovered from 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.