Journal of Hydrology: Regional Studies 27 (2020) 100652

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: Methodology and case studies in Andean watersheds

Juan F. Farfán[a,*], Karina Palacios[b], Jacinto Ulloa[c], Alex Avilés[b]

[a] University of A Coruña, Water and Environmental Engineering Group, Elviña Campus, n, 15071 A Coruña, Spain
[b] Environmental Engineering Career, Faculty of Chemical Sciences, University of Cuenca, Av. 12 de abril, s/n, Av. Loja, 010203, Cuenca, Ecuador
[c] KU Leuven, Department of Civil Engineering, Kasteelpark Arenberg 40, B-3001 Leuven, Belgium

ARTICLE INFO

Keywords:
Neural networks
Hydrological models
Flow forecasting
Andean watersheds
Ecuador

ABSTRACT

Study region: The present study was conducted in the Machángara Alto and Chulco rivers, which belong to the Paute basin in the provinces of Azuay and Cañar in southern Ecuador.
Study focus: Andean watersheds are important providers of water supply for human consumption, food supply, energy generation, industrial water use, and ecosystem services and functions for many cities in Ecuador and in the rest of South America. In these regions, accurate quantification and prediction of water flow is challenging, mainly due to significant climatic variability and sparse monitoring networks. In the context of flow forecasting, this work evaluates the accuracy of two physical models (WEAP and GR2M) and two models based on artificial neural networks (ANN) that use meteorological data as input variables. Then, a hybrid technique is proposed, using the time series generated by the individual models as inputs of a new ANN. This approach aims to increase the accuracy of the simulated flow by combining and exploiting the information provided by physical and data-driven models. To assess the performance of the proposed methodology, statistical analyses are conducted for two case studies in the Andean region, where comparative analyses are performed for the individual models and the hybrid technique.
New hydrological insights: The results indicate that the proposed technique is able to improve the individual performance of physical and ANN-based models, yielding good results in the calibration and validation stages for the two case studies. Specifically, increases in NSE were observed from 0.64 to 0.99 in the MachÁngara Alto river, and from 0.88 to 0.99 in the Chulco river. Higher accuracy of the hybrid technique was observed for all evaluation criteria considered in the analyses. The performance of the hybrid technique was also reflected in terms of water supply and demand, suggesting possible applications for the regional management of water resources, where accurate flow predictions are of utmost importance.

1. Introduction

Andean mountains harbor a unique ecosystem known as *páramo* (Flores-López et al., 2016). Currently, in the countries of the Andean region, it is estimated that 40 million people depend directly on these ecosystems as their main sources of water (Josse et al.,

* Corresponding author.
E-mail addresses: juan.farfand@udc.es (J.F. Farfán), jacintoisrael.ulloa@kuleuven.be (J. Ulloa).

https://doi.org/10.1016/j.ejrh.2019.100652
Received 25 November 2018; Received in revised form 14 November 2019; Accepted 2 December 2019

2011). Nevertheless, due to the geophysical complexity of the *páramo*, its hydrological processes remain poorly studied (Buytaert et al., 2011; Flores-López et al., 2016). The use of numerical models may be a viable tool to tackle such difficulties, and to provide crucial information for the regional management of water resources. For instance, the accuracy of flow predictions generated from different models can directly affect the planning of the operation of reservoirs, used for the generation of hydroelectric energy, irrigation and/or human consumption (Jiang et al., 2016; Veintimilla-Reyes et al., 2016; Wang et al., 2017).

Hydrological modeling can be divided into two categories: physical methods and data-driven methods (Weimin et al., 2013). Physical approaches typically involve knowledge-based conceptual components, which provide significant hydrological insights (Hosseini and Mahjouri, 2016). The main disadvantage of these modeling techniques is the presence of a large number of regionally dependent parameters that require calibration, making their optimization difficult (Si-min et al., 2009). On the other hand, the flow prediction of data-driven methods is based on the statistical relationship between hydrological variables (Vrugt et al., 2003). These techniques are capable of capturing patterns in extensive data sets, based on purely numerical input-output relationships. Consequently, data-driven methods have been widely used in hydrological modeling to describe rainfall-runoff processes (Mu noz et al., 2018). Despite the significant predictive skills of data-driven methods, their black-box nature lacks hydrological significance and robustness (Wang et al., 2017). For instance, failure to identify the driving factors of rainfall-runoff processes strongly limits flow predictions in ungauged watersheds, where models based on observable physical parameters may provide insightful information.

This study explores two physical models for flow simulation: WEAP and GR2M. The WEAP model (developed by the Stockholm Environment Institute) is based on water balance at a monthly time scale, where the user represents the system of water resources based on supply (rivers, groundwater and reservoirs), water demand and ecosystem requirements (Mounir et al., 2011). This model has shown success in studies of water supply and demand in several river basins, as well as in the assessment of climate change impacts (Al-Omari et al., 2014; Yi and Sandoval Solis, 2015; Abdella et al., 2017; Singh, 2018). Concerning regions relatively close to the present study area, the WEAP model was used by Lema Changoluisa and Plaza Quezada (2009) to assess climate change scenarios for the Pastaza river basin in Ecuador. The GR2M is a model based on parameter calibration that was proposed by the Center for Agricultural Research and Environmental Engineering of France for the prediction of flows using precipitation and evapotranspiration time series at a monthly scale (Hernández et al., 2013). The parsimony of the model guarantees robustness, which allows for several applications in different countries and regions, including the assessment of climate change (Makhlouf and Michel, 1994; Zolfaghari et al., 2013; Lavado Casimiro et al., 2011). Moreover, the model has been used in conjunction with other methodologies to improve its results (Ytoui, 2014; Huard and Mailhot, 2008). Regarding specific regions, Hernández et al. (2013) performed the calibration and validation of the GR2M in the upper basin of the Nazas river in Mexico. This river serves as a source of supply for the LÁzaro CÁrdenas and Francisco Zarco dams, which provide water for the irrigation of 170,000 ha in the area. In Peru, Alcántara et al. (2014) reported positive results from the GR2M in the Jequetepeque river.

Artificial neural networks (ANN) are computing systems that represent certain behavioral characteristics of the human brain (Li et al., 2018). They have been successfully used as data-driven prediction methods in reservoir operation and management of water resources (Maier and Dandy, 2000; Aqil et al., 2007; Iliadis and Maris, 2007; Wang et al., 2017). In relation to the present study area, Veintimilla-Reyes et al. (2016) implemented an ANN model for real-time flow prediction in the Tomebamba river in the city of Cuenca, Ecuador. In other regions, Wang et al. (2017) proposed a back-propagation neural network algorithm and applied it to the semi-distributed Xinanjiang model at an hourly time scale. The authors evaluated the procedure in the Dingan River in China and reported significant improvements. Another hybrid approach was explored in the upper Yalongjiang River Basin in China by Li et al. (2018). The authors evaluated the SWAT, VIC and BTOPMC models in independent simulations and combined the results in a new ANN. The results indicate that the proposed methodology improves the individual results of the models at a daily time scale.

The objective of this study is to implement and evaluate a hybrid technique that uses the monthly simulated flow of physical and ANN-based models as input variables, and configures a new ANN with observed flow as the target variable. The proposed technique aims to overcome the individual limitations of physical and data-driven approaches by combining the robustness and hydrological significance of physical models with the black-box predicting skills of data-driven methods. To this end, we employ two physical models (WEAP and GR2M) and two ANN-based models with meteorological input variables (precipitation and evapotranspiration). The individual performances are analyzed and compared with the results of the hybrid technique using several evaluation criteria, including a supply-demand analysis for the region of interest. The study is conducted for two rivers in the Ecuadorian Tropical Andes, where the complex climatic variability of the region implies significant challenges in flow forecasting.

2. Study area and data

The study area encompasses the Machángara Alto and Chulco rivers, which cover areas of 132 km^2 and 66 km^2, respectively. Both rivers are located in the Machángara sub-basin of the Paute river basin, which is located in the provinces of Azuay and Cañar in southern Ecuador (Fig. 1). The Machángara sub-basin supplies water for two reservoirs: Chanlud and El Labrado, with storage capacities of 16.25 hm^3 and 6.25 hm^3, respectively. These reservoirs supply water for irrigation and human consumption.

This region is influenced by air masses from the Pacific ocean and the Amazon basin, and by complex climatic factors including the movement of the Intertropical Convergence Zone (ITCZ), the relief of orographic rainfall, and marine currents from the Pacific Ocean, particularly, Humboldt and El Niño (Celleri et al., 2007; Coltorti and Ollier, 2000; Pourrut and Gómez, 1998; Sarmiento, 1986). As shown by Celleri et al. (2007), precipitation in the Machángara páramos presents a fairly uniform spatial distribution, while seasonality is virtually non-existent. Average rainfall, average temperature and relative humidity range from 856 mm to 1309 mm, from 8.1 °C to 14 °C, and from 76% to 88%, respectively. The altitude of the Machángara river basin ranges from 2500 to 4500 m.a.s.l., where topographic conditions play a significant role in meteorological processes (Estrella and Tobar, 2013).

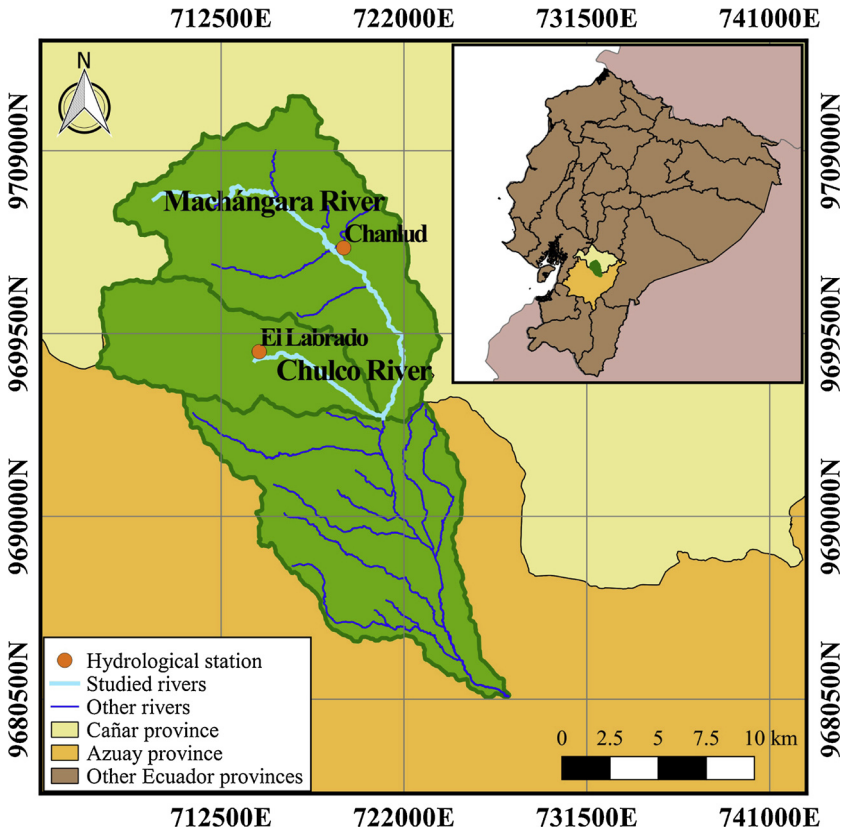

Fig. 1. Location of the study area and the hydrological stations.

This work covers a time period from January 1979 to December 2010, for which data is widely available from the Chanlud and El Labrado stations, located in the Machángara river and the Chulco river, respectively (Fig. 1). Consequently, both stations were selected for the analysis. These meteorological stations are maintained by the National Institute of Meteorology and Hydrology of Ecuador (INHAMI). A summary of the meteorological data for both stations is provided in Tables 1 and 2.

3. Methodology

This section starts with the definition of the calibration and validation periods. Then, the four models adopted for the analyses are presented: (1) the WEAP physical model, (2) the GR2M physical model, (3) ANN with precipitation data entry (hereafter P-NN) and (4) ANN with precipitation and evapotranspiration data entry (hereafter P-ET_0-NN). The presentation of the hybrid technique follows, which intends to increase the predictive skills of the four individual models. Criteria to evaluate the goodness of fit of the different approaches are explained, followed by details on the validation procedure.

3.1. Calibration and validation periods

The observed flow series were divided into two statistically similar periods: 75% of the data was selected as a calibration period, while the remaining 25% was used for validation. To evaluate statistical similarity, the presence of seasonal trends was first assessed using a Mann–Kendall test (Mann, 1945; Kendall, 1975) for each river, using the computer tools developed by McLeod (2005). For the

Table 1
Summary of meteorological data for the Chanlud station in the Machángara river.

Variable	Minimum	Q_1	Median	Mean	Q_3	Maximum
Streamflow (hm^3/month)	0.49	3.56	5.767	6.74	9.04	20.399
Precipitation (mm/month)	21.20	89.35	12.17	121.39	149.650	306.10
Potential evapotranspiration (mm/month)	35.84	44.37	48.31	47.59	50.85	56.30
Temperature (°C)	6.01	7.76	8.49	8.33	8.90	10.24
Relative humidity (%)	74.00	88.00	90.00	89.29	90.00	97.00

Table 2
Summary of meteorological data for the El Labrado station in the Chulco river.

Variable	Minimum	Q_1	Median	Mean	Q_3	Maximum
Streamflow (hm^3/month)	0.2	1.68	2.80	3.24	4.40	9.90
Precipitation (mm/month)	12.80	68.47	97.39	105.00	135.70	287.30
Potential evapotranspiration (mm/month)	35.38	44.85	49.43	48.20	51.52	58.15
Temperature (°C)	6.20	8.24	8.98	8.782	9.40	10.80
Relative humidity (%)	79.00	87.00	90.00	89.420	91.250	98.00

Machángara river, *p*-values of 0.071 for the calibration period and 0.79 for the validation period were obtained, while the Chulco river presented *p*-values of 0.06 for the calibration period and 0.77 for the validation period. In both cases, the *p*-values larger than 0.05 indicate no significant trends. The box plots in Figs. 2 and 3show the monthly flow distribution for both periods, which indicate similar temporal variability. A statistical analysis for both study periods is summarized in Tables 3 and 4.

3.2. WEAP

The WEAP physical model is a useful tool for the integrated planning of water resources. It requires a calibration phase of input variables such as monthly precipitation, monthly average temperature, monthly relative humidity, monthly wind speed and monthly flow. The calibration parameters consist of water storage capacity in the root zone, water storage capacity in the deep zone, runoff resistance factor and root zone conductivity (Al-Omari et al., 2014; Yi and Sandoval Solis, 2015). For hydrological modeling, the model adopts the Soil Moisture Model method. Its operation is based on the representation of water balance in two layers of soil: a higher one, in which evapotranspiration is simulated considering rainfall and irrigation, surface runoff, subsurface runoff and deep percolation, and a lower one, that simulates the passage of the base flow and changes in the water content of the soil (Al-Omari et al., 2014; Yi and Sandoval Solis, 2015). The WEAP model is based on a semi-distributed approach, which divides the basin into several fractions of area, each of which can represent different soil types or land uses. Water balance is then performed for each fraction. Further details on the model are presented by Al-Omari et al. (2014) and Yi and Sandoval Solis (2015).

The present calibration and validation of the WEAP model employ data available in the period from 1979 to 2010. The calibration phase was carried out with data between 1979 and 1999, while validation was performed between 2000 and 2010. The model was optimized by means of a trial-and-error procedure. As such, the performance of the WEAP model is associated with this specific calibrated product.

3.3. GR2M

GR2M is a conceptual physical model that estimates flows from rainfall and evapotranspiration data. Its calibration requires monthly precipitation, monthly mean evapotranspiration and monthly flow as input variables. The model operation is based on two

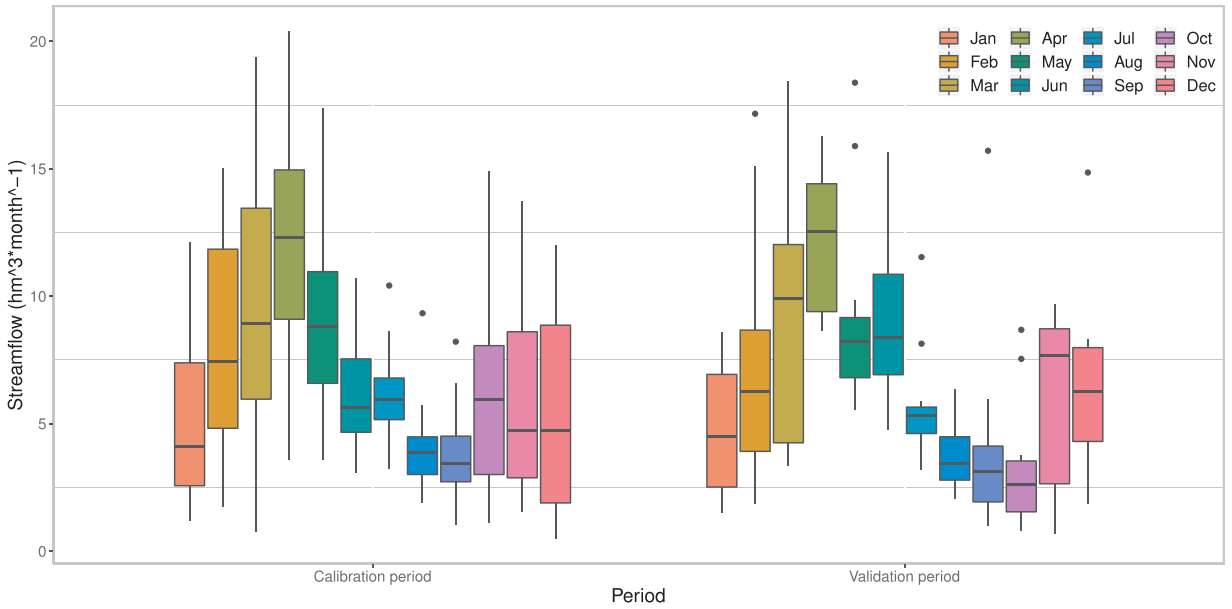

Fig. 2. Box plot of monthly flow distribution in the calibration and validation periods for the Machángara river.

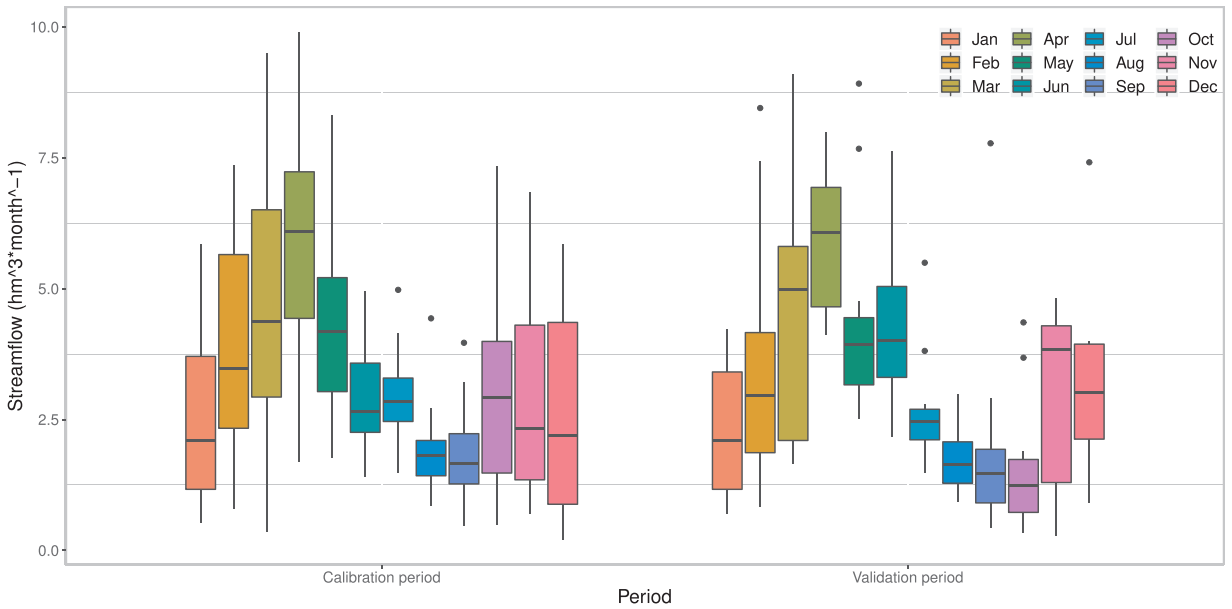

Fig. 3. Box plot of monthly flow distribution in the calibration and validation periods for the Chulco river.

Table 3
Statistical summary of the data in the calibration and validation periods for the Machángara river.

Stage	Minimum	Q_1	Median	Mean	Q_3	Maximum	Standard dev.
Calibration	0.49	3.63	5.64	6.73	9.10	20.39	4.09
Validation	0.70	3.34	6.09	6.75	8.88	18.45	4.23

Table 4
Statistical summary of the data in the calibration and validation periods for the Chulco river.

Stage	Minimum	Q_1	Median	Mean	Q_3	Maximum	Standard dev.
Calibration	0.21	1.76	2.71	3.24	4.41	9.90	2.00
Validation	0.28	1.58	2.90	3.25	4.36	9.10	2.08

tanks: the reservoir-soil, related to a production function, and the gravitational water reservoir, where the contribution is instantaneous at the beginning and gradually empties with the passage of time. The release of the gravitational water reservoir depends on the storage level (Mouelhi et al., 2006). Potential evapotranspiration is computed as a function of average temperature, following the method proposed by Thornthwaite et al. (1954). For the latest version of the GR2M model, the reader is referred to Mouelhi (2003).

The GR2M model consists of two parameters: the tank capacity and a dimensionless constant. These parameters are adjusted by optimizing an objective function, for which the Nash–Sutcliffe coefficient (NSE) is adopted (Mouelhi, 2003). In order to achieve an optimal set of parameters, the Generalized Gradient Code for Nonlinear Programming developed by Lasdon et al. (1976) was used in this study. As for the WEAP model, the performance of GR2M is hereby associated with this specific calibrated product.

3.4. ANN-based models: P-NN and P-ET$_0$-NN

In ANN-based techniques, the architecture of the neuron is represented as the sum of external stimuli z_j, followed by a non-linear activation function $y_j = f(z_j)$. This function determines the output activity of the neuron. This type of ANN architecture is known as the McCulloch-Pitts perceptron (Tkacz et al., 1999; Fidalgo et al., 2013). Therein, the user enters the data for both input and output variables to train the ANN. Through a set of random weights, the network begins the learning process in search for an optimal set, which allows for pattern classification and error minimization (Crone, 2002; Perea et al., 2016).

The present study uses a feed-forward back-propagation neural network algorithm available in the MATLAB toolbox NNtool (Zhong and Xie, 2012). The number of neurons in the hidden layer that yields an optimal performance was estimated by means of a trial-and-error procedure. The hyperbolic sigmoidal tangent was used as a transfer function. For training, the Bayesian regularization algorithm was applied, which minimizes a convex combination of the mean-square-error. This configuration of the ANN is consistent

with similar studies that report positive results (e.g., Fidalgo et al., 2013; Perea et al., 2016).

Two ANN-based modeling alternatives are explored. In the first (P-NN), precipitation estimates comprised in the period from 1979 to 1999 were supplied to the ANN as input data (independent variable), while the flow measurements in the same period were used as output data (dependent variable). In the second alternative (P-ET_0-NN), the P-NN model was enhanced by incorporating evapotranspiration input data, which was calculated by means of the Thornwaite method (Pereira and Pruitt, 2004). Both ANN models were subjected to a validation process with the corresponding input data in the period between 2000 and 2010.

3.4.1. ANN-based hybrid technique

The proposed hybrid technique is a multi-model approach that aims to increase accuracy in flow predictions. For this procedure, the 4 time series obtained from the previously described models are used as the entry of a new ANN. The time series of the ANN-based models (P-NN and P-ET_0-NN) correspond to the results of the optimal number of neurons. The configuration and training of the ANN follows the trial-and-error procedure and Bayesian regularization algorithm, as described for P-NN and P-ET_0-NN.

3.5. Goodness-of-fit analysis

A goodness-of-fit analysis was performed to compare the predicted time series to the observed data in both rivers. The comparisons allow to observe the capability of the different models to capture temporal variation and extreme values.

To avoid subjectivity in the analyses, a combination of graphical results, absolute value error and normalized goodness-of-fit statistics was performed, following the recommendations of Ritter and Muñoz-Carpena (2013), where NSE is used to indicate goodness-of-fit. Moreover, following McCuen et al. (2006), NSE is complemented by bias quantification. Table 5 presents a summary of the threshold criteria for NSE and PBIAS, which are based on Ritter and Muñoz-Carpena (2013) and Moriasi et al. (2007), respectively. This information is further complemented by Root Mean Squared Error to analyze the deviation of simulated flows.

3.6. Application to water resources management

To evaluate the ability to integrate the present technique as a tool for the management of water resources in Andean regions, the WEAP supply-demand component (Yates et al., 2005) was used to perform a water supply-demand analysis for the study area. The water supply in hm^3/month was computed from the simulated flow of the individual models and the hybrid technique in both rivers. These quantities were then compared to the demand, which was quantified as the sum of 3 demand points in the Machángara catchment that are supplied by the Chulco and Machángara rivers. The demand factors consist of water for purification in the Tixán treatment plant and the irrigation channels Machángara and La Dolorosa. These demand values are available in ETAPA (2016) and Estrella and Tobar (2013). For both reservoirs, the inflows are measured through the estimation of storage volumes with respect to time.

The distribution model was constructed using the WEAP supply-demand component, in which water balance was performed. A scheme of the system is shown in Fig. 4, showing the reservoirs, channels, diversions and demands.

4. Results

4.1. Machángara river

Fig. 5 shows the observed flow for the Chanlud station in the Machangara river and the simulated flow for the WEAP, GR2M, P-NN and P-ET_0-NN models in both calibration and validation phases. The plots reveal that while WEAP generates a relatively good fit, it is the only model that overestimates both high and low extreme values. On the other hand, the GR2M model adequately represents the dynamics of the flow, but shows a poor estimation of peak flows and overestimates low flows.[1] In turn, P-NN and P-ET_0-NN present a relatively good fit, but fail to capture the lowest flow rates in the calibration and validation phases. In addition, high flows are underestimated.

Fig. 6 shows the observed flow rates on the horizontal axis, the simulated on the vertical axis, and a 1:1 line to further visualize the accuracy of the predictions. The flows simulated by the WEAP model are above the 1:1 line, indicating bias, particular for high flow rates. For this model, the calibration values for both NSE and PBIAS are unsatisfactory (Table 5), with PBIAS indicating a significant overestimation of 34.67%. An RMSE value of 2.47 hm^3/month was obtained, corresponding to the highest of all 4 models. In the validation stage, NSE increases to 0.72, corresponding to satisfactory, while PBIAS and RMSE decrease to 31.07% (unsatisfactory) and 2.25 hm^3/month, respectively. As in the calibration stage, bias increases for high flow rates, with possible outliers. In the case of the GR2M model, a more dispersed distribution is observed around the 1:1 line in the calibration stage, where the values below the 1:1 line indicate underestimation. However, in some cases, higher overestimations than in the WEAP model are observed. A satisfactory NSE of 0.72 was obtained, with a very good PBIAS of 4.15%. However, a high RMSE of 2.17 hm^3/month was obtained. In the validation stage, NSE decreases to an unsatisfactory value of 0.60, which can be attributed to a PBIAS shift to underestimation, and an increase in RMSE to 2.66 hm^3/month.

P-NN shows a calibration satisfactory NSE value of 0.75, a very good PBIAS of 0.29% and an RMSE of 2.05 hm^3/month. The

[1] We recall that the results of the WEAP and GR2M models reflect the performance of the specific calibrated products used in the present study.

Table 5
Goodness-of-fit values for performance analysis.

Goodness-of-fit	NSE	PBIAS
Very good	NSE ≥ 0.90	|PBIAS| ≤ 10 %
Goodness of fit	0.80 ≤ NSE < 0.90	10% ≤ |PBIAS| < 15 %
Satisfactory	0.65 ≤ NSE < 0.80	15% ≤ |PBIAS| ≤ 25 %
Unsatisfactory	NSE < 0.65	|PBIAS| ≥ 10 %

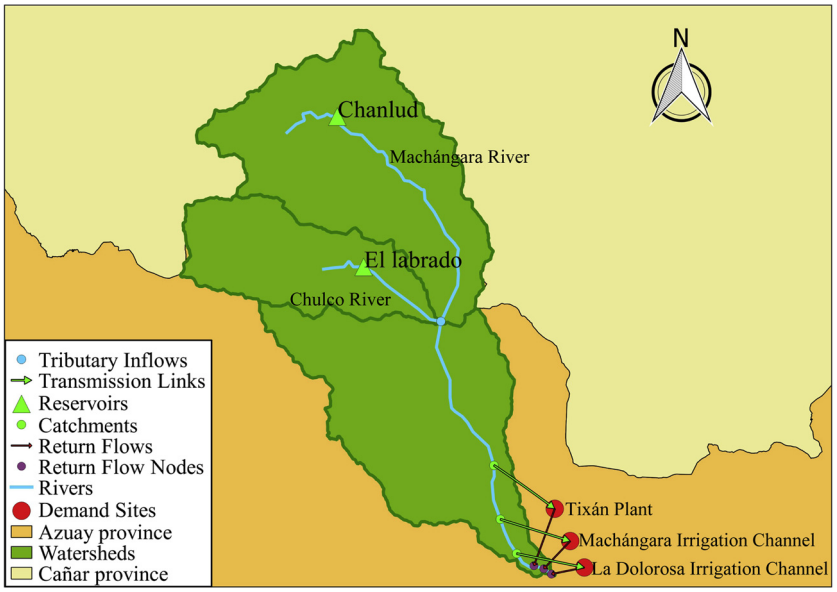

Fig. 4. WEAP Water Resources system model.

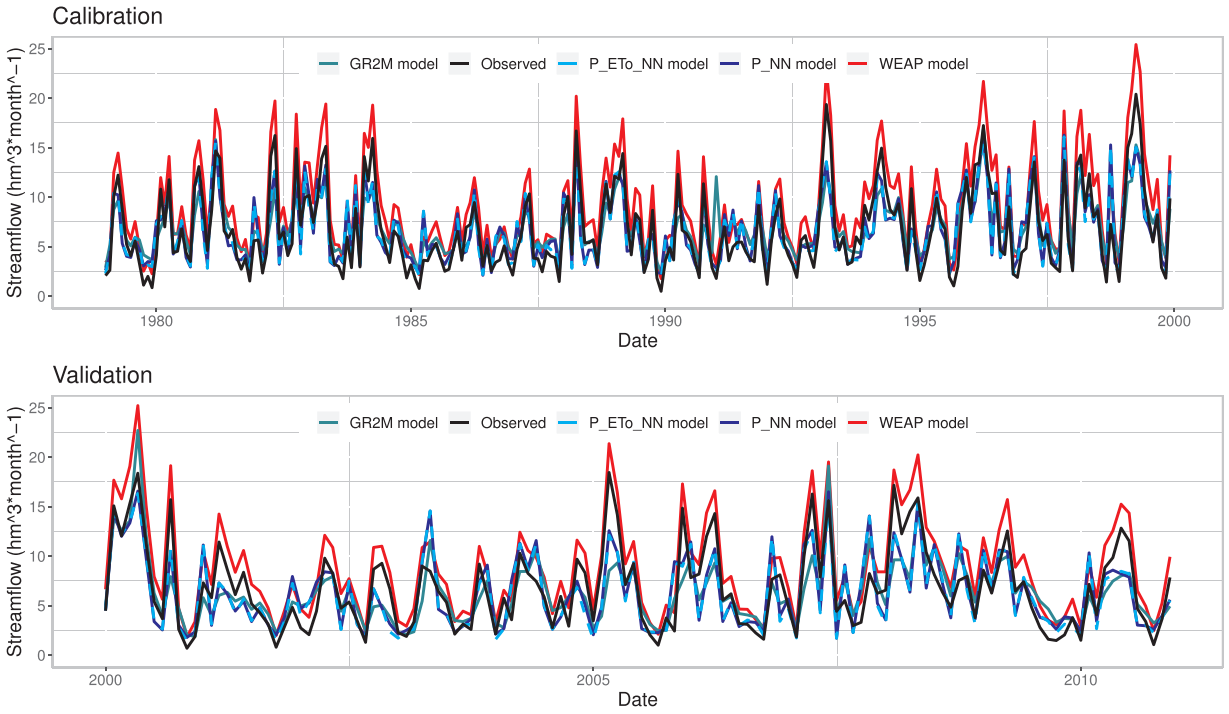

Fig. 5. Hydrographs comparing the observed flow in the Machángara river to the simulated flow of the individual models in the calibration (top) and validation (bottom) stages.

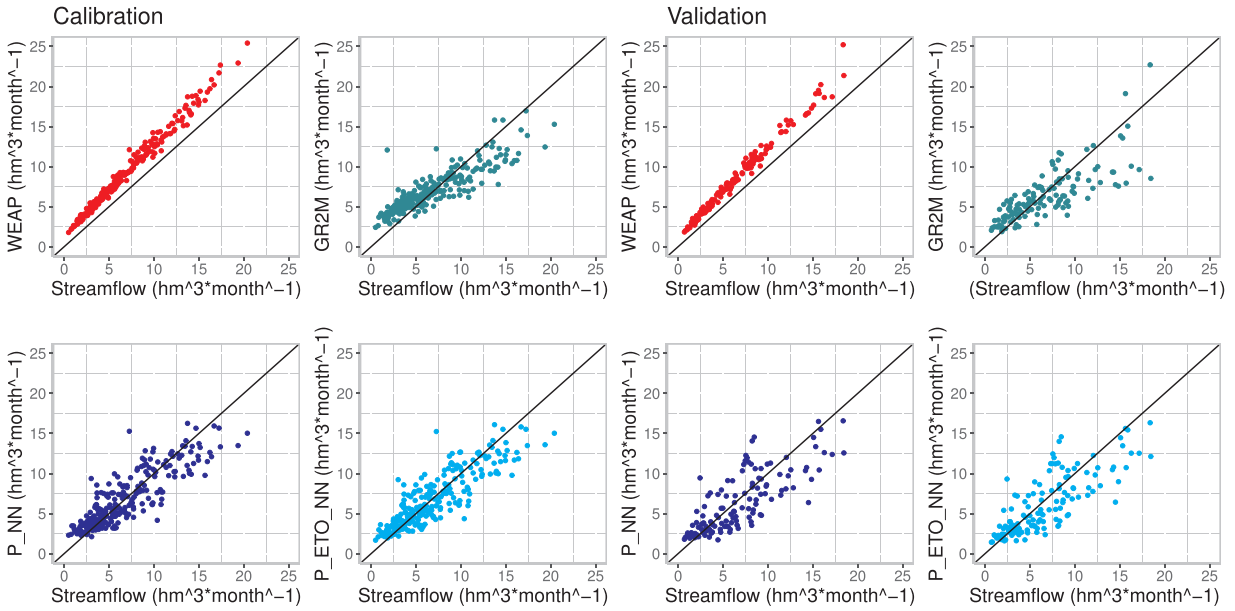

Fig. 6. Observed streamflow vs simulated streamflow in the calibration (left) and validation (right) stages for the Machángara river.

results indicate no significant outliers and a very low model bias, showing rather acceptable results. In the validation stage, NSE decreases to 0.61, accompanied by an underestimation with a PBIAS of −6.23%. In the calibration stage, the P-ET_0-NN model presents a satisfactory NSE of 0.75 and a very good PBIAS of 1.16%, with an RMSE of 2.02 hm^3/month. However, the validation results for this model show an unsatisfactory NSE of 0.61, accompanied by underestimation with a PBIAS of −8.11%.

Fig. 7 shows the results of the proposed hybrid technique in the calibration and validation phases. The time series indicate that the method accurately captures the flow dynamics of the Machángara river, drastically improving the results of the individual models shown in Fig. 5.

Fig. 8 shows the data closely aligned to the 1:1 line, indicating significant agreement with the observed data. This performance

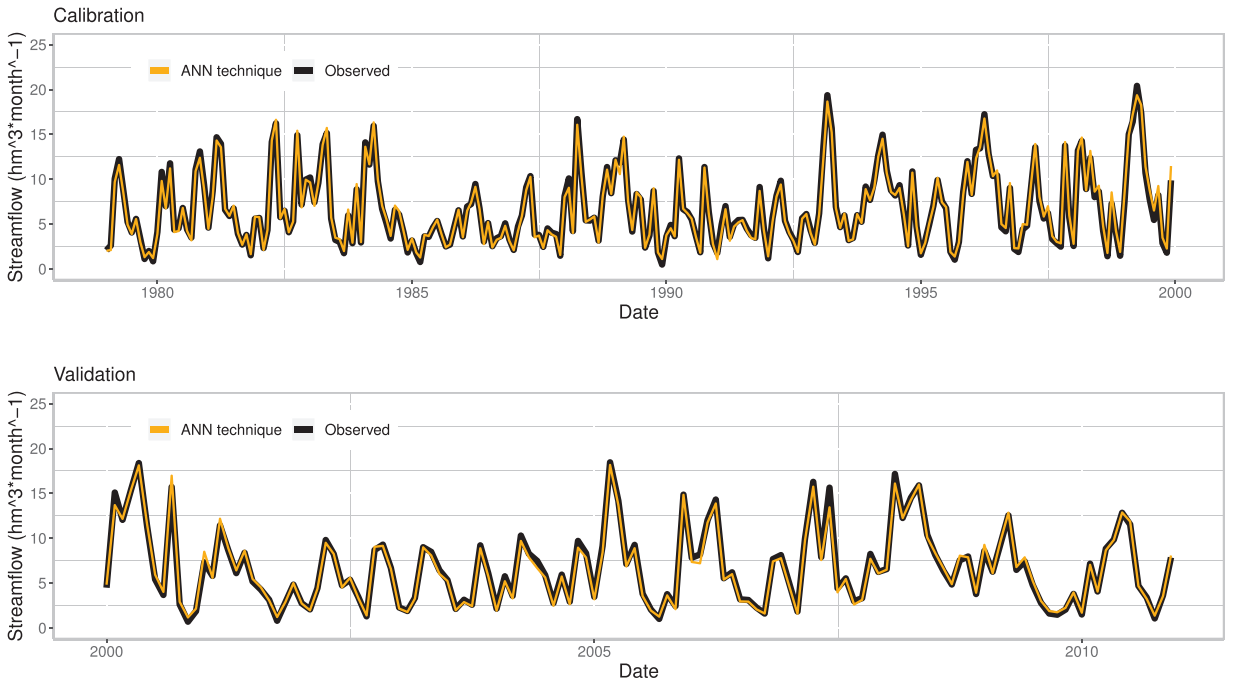

Fig. 7. Hydrographs comparing the observed flow in the Machángara river to the simulated flow of the hybrid technique in the calibration (top) and validation (bottom) stages.

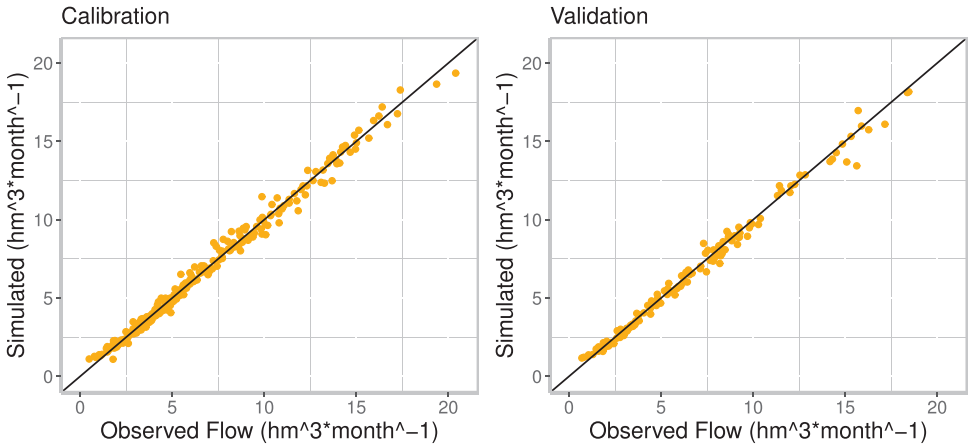

Fig. 8. Observed streamflow vs simulated streamflow of the hybrid technique in the calibration (left) and validation (right) stages for the Machángara river.

remains consistent in the entire range of flow rates in the calibration stage, with no significant outliers or non-linear trends. A very good NSE of 0.991 is obtained, with a very good PBIAS of 0.24%. In the validation stage, the results show an improvement with respect to all individual models, yielding an NSE of 0.99. Moreover, the persistent underestimation in the individual models (apart from WEAP) significantly decreases to a PBIAS of −1.33%. The RMSE is also remarkably low, at 0.42 hm^3/month.

Table 6 presents the results of the 4 individual models in the Machángara river. For the calibration phase, the NSE coefficient corresponds to a very good fit in all cases. Both, P-NN and P-ET_0-NN models present the best results in terms of NSE, with values of 0.75. WEAP model presents the lowest NSE and an unsatisfactory PBIAS of 34.67%. The P-NN and P-ET_0-NN models present a good fit in all coefficients. In the validation phase, WEAP presents the best results, yielding an acceptable fit in terms of NSE. However, the value of PBIAS was unsatisfactory, presenting the lowest performance among the 4 models. GR2M shows a bad NSE equal to 0.60. Similar bad fit is obtained for both data-driven models.

4.2. Chulco river

Fig. 9 shows the observed flow of the El Labrado station in the Chulco river and the simulated of the WEAP, GR2M, P-NN and P-ET_0-NN models for both calibration and validation phases. The plots reveal that while the calibrated WEAP product presents a good fit, it is the only model that overestimates both high and low extremes values. On the other hand, the product calibrated for the GR2M model adequately represents the flow dynamics, but shows a poor estimation of peak flows and overestimates low flows. Finally, the P-NN and P-ET_0-NN present a relatively good fit, but fail to capture the lowest flow rates in the calibration and validation phases. In addition, high flows are underestimated.

Fig. 10 shows the observed flow rates on the horizontal axis, the simulated on the vertical axis, and a 1:1 line to further visualize the accuracy of the predictions. The flows simulated by the WEAP model are above the 1:1 line, indicating bias, particular for high flow rates. For this model, the calibration values for both NSE and PBIAS are unsatisfactory (Table 7), with PBIAS indicating a significant overestimation of 19.39%. In addition, an RMSE value of 0.8 hm^3/month was obtained. In the validation stage, NSE decreases to 0.87, corresponding to good, while PBIAS increases to 31.07 hm^3/month (unsatisfactory) and RMSE decreases 0.76hm^3/month. As in the calibration stage, bias increases for high flow rates, with possible outliers. In the case of the GR2M model, a more dispersed distribution is observed around the 1:1 line in the calibration stage, where the values below the 1:1 line indicate underestimation. However, in some cases, higher overestimations than in the WEAP model are observed. A satisfactory NSE of 0.78 was obtained, with a very good PBIAS of 1.36% and an RMSE of 0.95 hm^3/month. In the validation stage, NSE increases to a good value of 0.80. Both PBIAS and RMSE decrease to 0.57 and 0.92 hm^3/month, respectively.

P-NN shows an acceptable calibration NSE value of 0.77, a very good PBIAS of −1.06% and an RMSE of 0.97 hm^3/month. The

Table 6
Calibration and validation results in the Machángara river.

Model	Calibration			Validation		
	NSE	PBIAS (%)	RMSE (hm^3/month)	NSE	PBIAS (%)	RMSE (hm^3/month)
WEAP	0.64	34.67	2.47	0.72	31.07	2.25
GR2M	0.72	4.16	2.17	0.60	−5.87	2.26
P-NN	0.75	0.29	2.05	0.61	−6.23	2.78
P-ET_0-NN	0.75	1.16	2.02	0.61	−8.11	2.75
Hybrid NN technique	0.99	0.24	0.40	0.99	−1.33	0.42

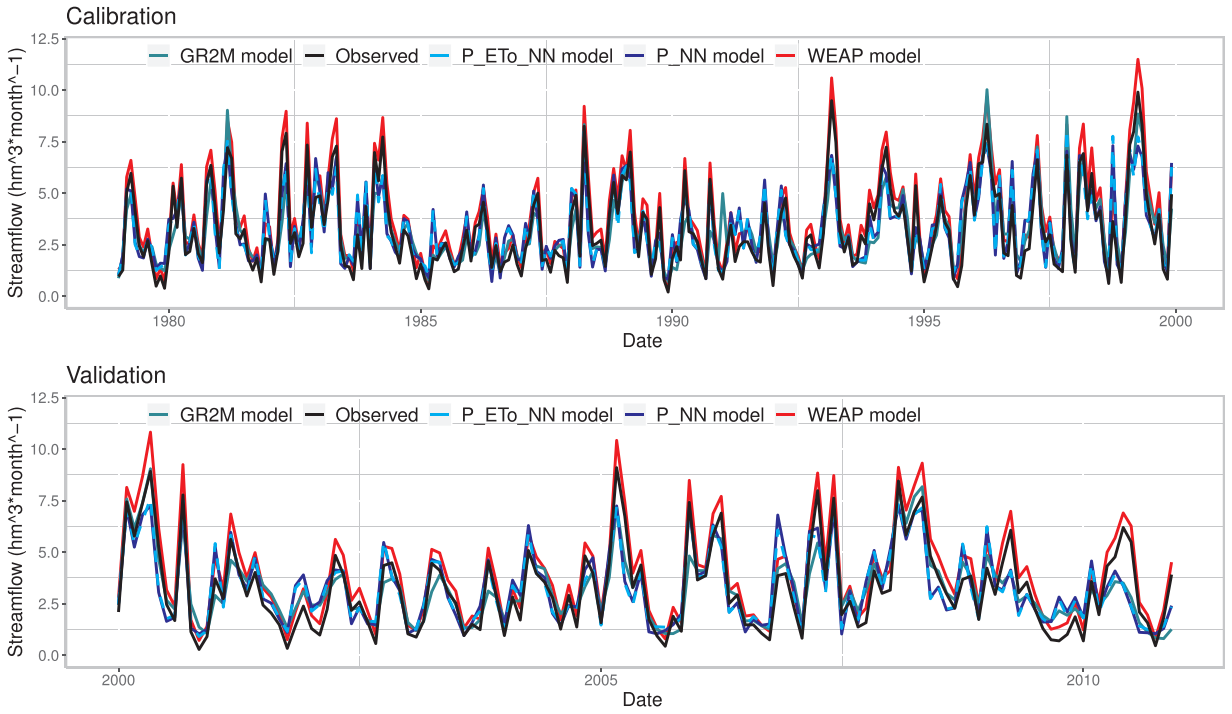

Fig. 9. Hydrographs comparing the observed flow in the Chulco river to the simulated flow of the individual models in the calibration (top) and validation (bottom) stages.

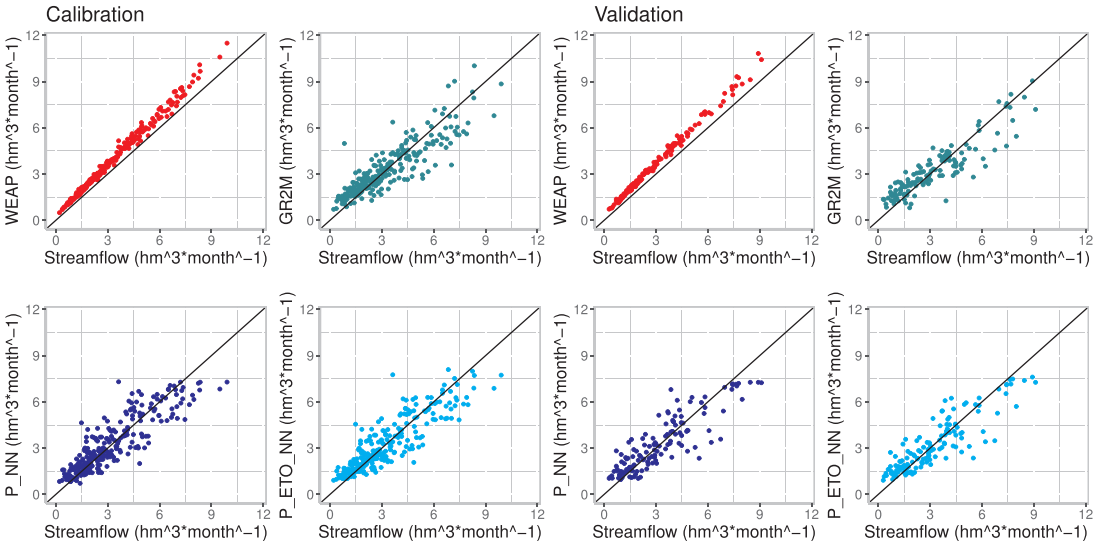

Fig. 10. Observed streamflow vs simulated streamflow in the calibration (left) and validation (right) stages for Chulco river.

results indicate no significant outliers and a very low model bias, showing rather acceptable results. In the validation stage, NSE decreases to 0.76, accompanied by an underestimation with a PBIAS of −1.36%. In the calibration stage, the P-ET_0-NN model presents an acceptable NSE of 0.77 and a very good PBIAS of 0.52%, with an RMSE of 0.96 hm^3/month. The validation results for this model show an acceptable NSE of 0.77, and a PBIAS of 0.98%. The results are summarized in Table 7.

Fig. 11 shows the results of the proposed hybrid technique in the calibration and validation phases. The time series indicate that the method accurately captures the flow dynamics of the Chulco river, drastically improving the results of the individual models shown in Table 7

Fig. 12 shows data closely aligned to the 1:1 line, indicating significant agreement with the observed data. This performance remains consistent in the entire range of flow rates in the calibration stage, with no significant outliers or non-linear trends. A very good NSE of 0.99 is obtained, with a very good PBIAS of 0.03%. In the validation stage, the results show an improvement with respect

Table 7
Calibration and validation results in Chulco river.

Model	Calibration			Validation		
	NSE	PBIAS (%)	RMSE (hm^3/month)	NSE	PBIAS (%)	RMSE (hm^3/month)
WEAP	0.88	19.39	0.80	0.87	21.80	0.76
GR2M	0.78	1.36	0.95	0.80	0.57	0.92
P-NN	0.77	−1.06	0.97	0.76	−1.36	1.02
P-ET_0-NN	0.77	0.52	0.96	0.77	0.31	0.98
Hybrid NN technique	0.99	0.03	0.10	0.99	2.61	0.19

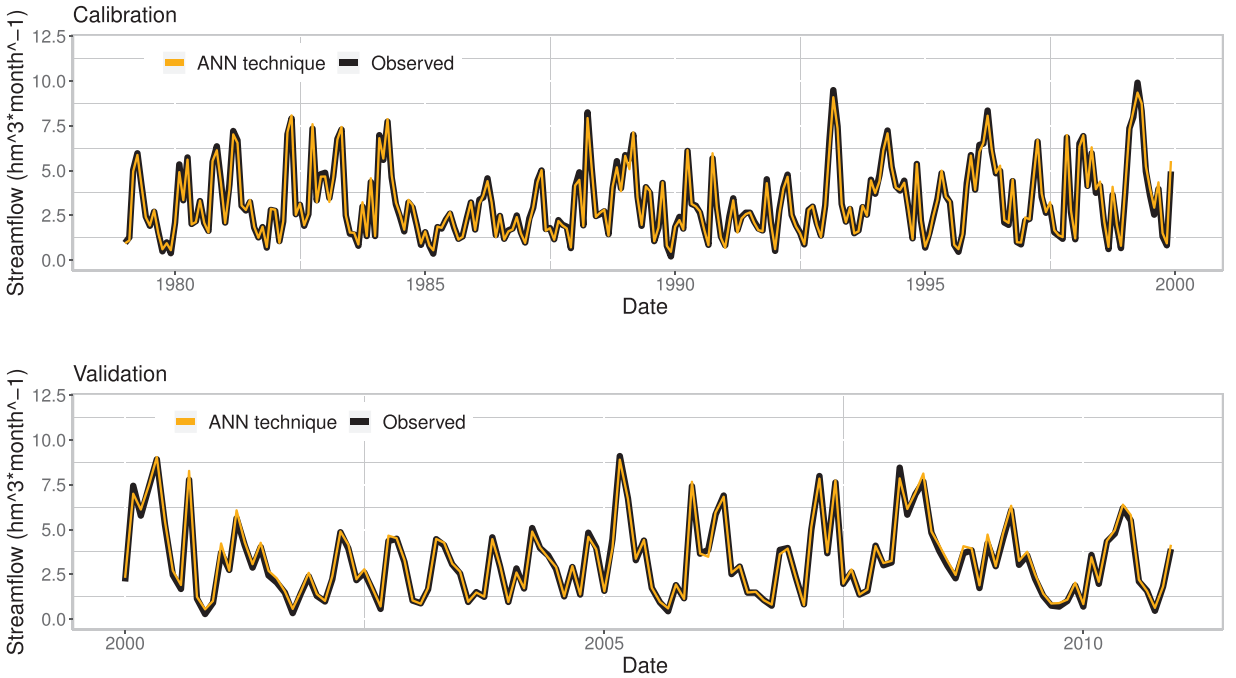

Fig. 11. Hydrographs comparing the observed flow in the Chulco river to the simulated flow of the hybrid technique in the calibration (top) and validation (bottom) stages.

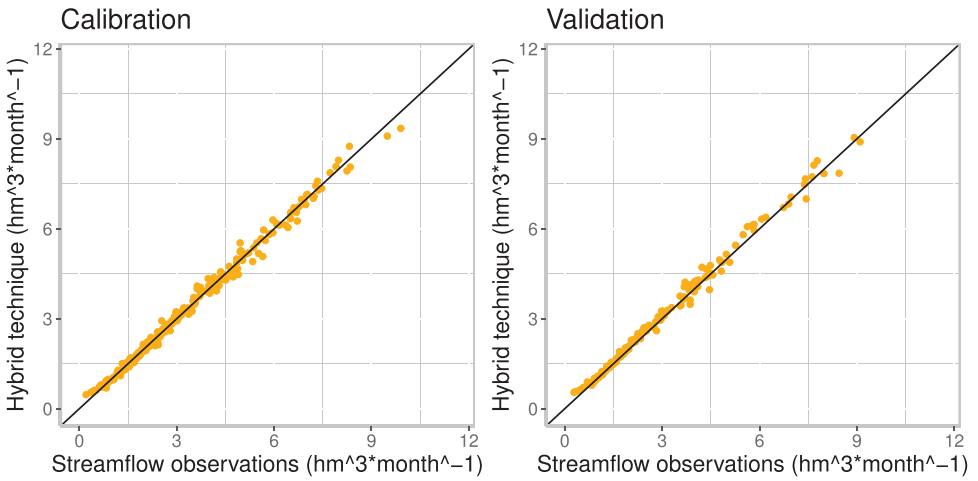

Fig. 12. Observed streamflow vs simulated streamflow of the hybrid technique in the calibration (left) and validation (right) stages for the Chulco river.

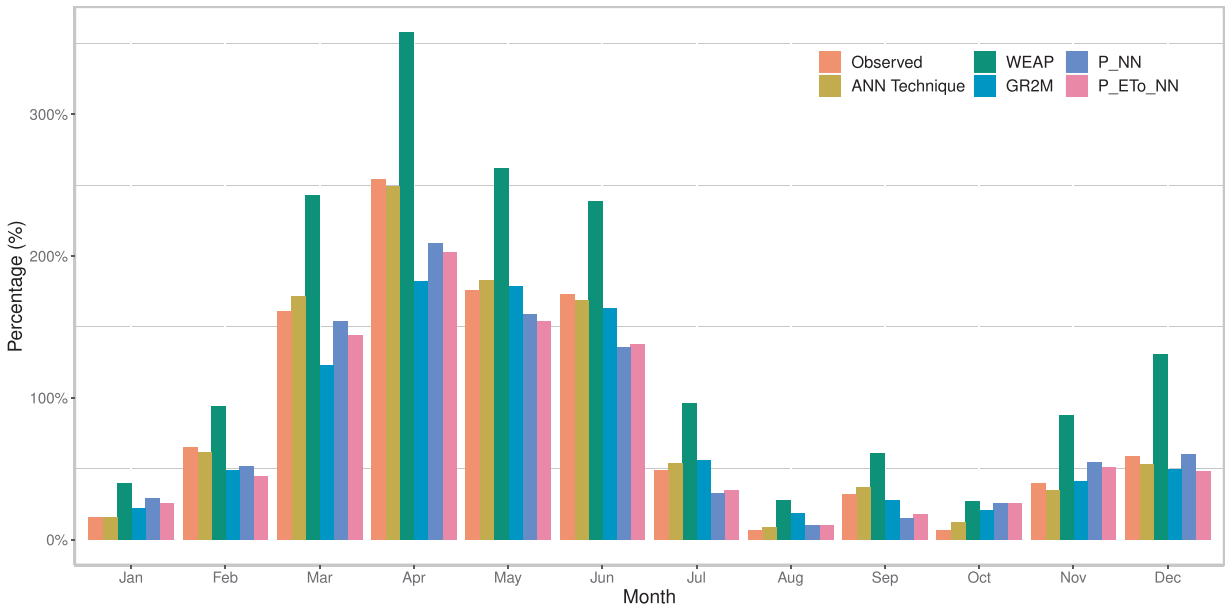

Fig. 13. Graphical representation of the results in the application of the individual models and the hybrid technique to a system of water resources.

to all individual models, yielding an NSE of 0.99. Moreover, the persistent underestimation in the individual models (apart from WEAP) significantly decreases to a PBIAS of 2.61%. The RMSE is also remarkably low, at 0.19 hm^3/month.

Table 7 presents the results of the 4 individual models in the Chulco river. For the calibration phase, the NSE coefficient corresponds to a very good fit in all cases. The WEAP model presents the best results in terms of NSE, with values of 0.88. However, a satisfactory PBIAS of 19.39% was obtained. The GR2M, P-NN and P-ET_0-NN models present a very good fit in all coefficients. In the validation phase, WEAP presents the best results, yielding a very good fit in terms of NSE. However, the value of PBIAS was satisfactory, presenting the lowest performance among the 4 models.

4.3. Application as a tool for water resources management

We compare the water supply of the individual techniques, the hybrid technique and the observed data to the downstream demand, obtained in terms of drinking water and irrigation in three demand sites (obtained from ETAPA, 2016; Estrella and Tobar, 2013). The results are shown in Fig. 13, which presents the surplus that was obtained in all cases. When compared to the observed data, the (calibrated) WEAP output supply presents a significant overestimation, which increases in the months of higher water supply. In the cases of March, April, May, and June, overestimations are observed with a difference of up to 100% (Table 8), in agreement with the PBIAS observed in both rivers. The results of GR2M, PNN and P-ET_0-NN deviate less than WEAP from the surplus of the observed data set. However, the results obtained by the GR2M model show an estimate of the surplus lower than all the models in wet months, such as April and May, while in most months with lower surpluses, such as January, August and October, high overestimations occurred. Both ANN models (PNN and P-ET_0-NN) show similar results, with underestimation from February to June. On the other hand, in the months of lower demand, i.e., from October to December, the simulations produced high overestimations.

Table 8
Percentage results of the surplus-deficit analysis for the individual models and the hybrid technique.

Month	Observed Surplus (+)	WEAP Surplus (+)	GR2M Surplus (+)	P-NN Surplus (+)	P-ET_0-NN Surplus (+)	Hybrid technique Surplus (+)
Jan	16%	40%	22%	29%	26%	16%
Feb	65%	94%	49%	52%	45%	62%
Mar	161%	243%	123%	154%	144%	172%
Apr	254%	358%	182%	209%	203%	249%
May	176%	262%	179%	159%	154%	183%
Jun	173%	239%	163%	136%	138%	169%
Jul	49%	96%	56%	33%	35%	54%
Aug	7%	28%	19%	10%	10%	9%
Sep	32%	61%	28%	15%	18%	37%
Oct	7%	27%	21%	26%	26%	12%
Nov	40%	88%	41%	55%	51%	35%
Dec	59%	131%	50%	60%	48%	53%

In some cases, such as October, the overestimation was almost as high as the results of WEAP for the same month. The hybrid technique presents very similar results to the observed surplus, with small deviations throughout the different months. However, as expected, all individual methods are consistently outperformed by the hybrid technique.

5. Discussion

Concerning the two physical models, we emphasize that the results reflect the performance of the present calibrated products, while improved calibration has not been studied. Nevertheless, the results of the WEAP-based product suggest that the accuracy of flow prediction using models that involve the calibration of several parameters is significantly compromised in regions with complex climatic conditions, as in the case of Andean watersheds. On the other hand, simpler models such as GR2M might not be able to represent the rainfall-runoff processes due to the lack of representative parameters. In turn, the ANN-based models proved to be more accurate tools for flow forecasting than the products derived from the physical models in both case studies.

Studies conducted by Li et al. (2018) in the Upper Yalongjiang Basin in China apply an alternative hybrid approach that resembles the present work. The authors use results from the SWAT, VIC and BTOPMC models as inputs and report very good results for simulations at a daily time scale, with NSE coefficients of 0.95 and 0.90 in the calibration and validation stages, respectively. Consequently Li et al. (2018) suggest that using the results of different models as the input of a neural network can increase the goodness of fit in the prediction of flows.

This hypothesis was verified for the hybrid technique proposed in the present study. Significant improvements with respect to all four individual models were observed in both case studies, with NSE values around 0.99 in both calibration and validation stages. To the knowledge of the authors, these are the most accurate results to date in the two catchments under study. This positive outcome may be attributed to the different nature of the information captured by the individual models, which is combined in the hybrid neural network. Both the WEAP and GR2M physical models embed information directly related to monthly runoff processes, while the P-NN and P-ET_0-NN models capture non-linear relationships between the input climatic data and the output flows.

To further assess the capabilities of the proposed technique, we compared the water supply of the individual models, the hybrid technique and the observed data to the downstream demand, obtained in terms of drinking water and irrigation in three demand sites. When compared to the observed data, as expected, all individual methods are consistently outperformed by the hybrid technique.

6. Conclusions

The objective of this study was to implement and evaluate a hybrid technique that uses the monthly simulated flow of two physical models (WEAP and GR2M) and two data-driven methods (with precipitation and reference evapotranspiration) as input variables, to configure a new ANN with observed streamflow as the target variable. The purpose of this technique was to overcome the limitations of physical and data-driven models by combining their individual capabilities. The results of the comparative statistical analyses indicate that the ANN-based hybrid technique provides major improvements over the individual models. Specifically, the ability to capture peak flows is enhanced, resulting in significantly better statistical values. The results also reveal that the hybrid technique is more consistent than the individual models, presenting similar results in the calibration and the validation phases.

From the present (positive) results, the conditions under which the hybrid technique fails to improve the results of physical and data-driven models could be not be addressed. Thus, additional case studies are required to reveal such conditions. In addition, the performance of the physical models (WEAP and GR2M) was studied using calibrated data, where the performance of the calibration procedure was not addressed. Consequently, studies with different data sets resulting from physical models with improved calibration techniques are required to better study the enhancement capabilities of the hybrid technique.

This work can provide the basis for future studies, for instance, on the performance of the hybrid technique for extrapolating predictions to ungauged regions. Moreover, ANN-based hybrid techniques for streamflow simulations emerge as a possible strategies to be used in the management of water resources, where accurate predictions are of utmost importance. In particular, accurate streamflow predictions are required to study the supply and demand of water resources in regions where surplus levels are close to the deficit threshold. For instance, considering that the population in the city of Cuenca grew from 104,470 to 329,928 inhabitants between 1974 and 2010, and that it is projected to reach around 861,682 by 2055 (Hermida et al., 2015), the scarce surplus in certain months (as shown in the present study for the months of January, August and October) implies concerns in terms of future availability. To address these cases, stochastic meteorological simulation methods (e.g., for precipitation and temperature) can be used to generate input data to be used in the hybrid technique. This could allow to assess water availability under different scenarios at a regional scale, accounting for the effects of population growth, climate change and new demand sites.

Conflict of interest

The authors declare that they have no conflict of interest.

Funding

This work was funded by the University of Cuenca through its Research Department (DIUC) as part of the project titled “Evaluación del riesgo de sequías en cuencas andinas reguladas influenciadas por la variabilidad climática y cambio climático. Caso

de estudio en la cuenca del río Machángara."

Acknowledgements

We appreciate the provision of information of the National Institute of Meteorology and Hydrology of Ecuador.

References

Abdella, E.J., Gosain, A.K., Khosa, R., 2017. Impact of climate change on irrigation and hydropower potential: a case of Upper Blue Nile Basin. AGU Fall Meeting Abstracts.
Al-Omari, A., Salman, A., Karablieh, E., 2014. The Red Dead Canal project: an adaptation option to climate change in Jordan. Desalin. Water Treat. 52 (13–15), 2833–2840.
Alcántara, A., Montalvo, N., Mejía, A., Ingol, E., 2014. Validación de modelos hidrológicos lluvia-escorrentía para su aplicación a la cabecera de cuenca del río jequetepeque. Revista del Instituto de Investigación de la Facultad de Ingeniería Geológica, Minera, Metalurgica y Geográfica 17 (33).
Aqil, M., Kita, I., Yano, A., Nishiyama, S., 2007. A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. J. Hydrol. 337 (1–2), 22–34.
Buytaert, W., Cuesta-Camacho, F., Tobón, C., 2011. Potential impacts of climate change on the environmental services of humid tropical alpine regions. Glob. Ecol. Biogeogr. 20 (1), 19–33.
Celleri, R., Willems, P., Buytaert, W., Feyen, J., 2007. Space-time rainfall variability in the Paute basin, Ecuadorian Andes. Hydrol. Process. 21 (24), 3316–3327.
Coltorti, M., Ollier, C., 2000. Geomorphic and tectonic evolution of the Ecuadorian Andes. Geomorphology 32 (1–2), 1–19.
Crone, S.F., 2002. Training artificial neural networks for time series prediction using asymmetric cost functions. Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02, vol. 5 2374–2380.
Estrella, R., Tobar, V., 2013. Hidrologia y Climatologia – Formulación del Plan de Manejo Integral de la Subcuenca del Río Machangara; ACOTECNIC Cia. Ltda. Consejo de Cuenca del Río Machangara, Cuenca, Ecuador.
ETAPA, 2016. Ampliación de la Capacodad de la Planta de Tratamiento de Agua Potable de Tixán: Módulo II. Subgerencia de Ingeniería de Proyectos.
Fidalgo, J.E.L.C., La Red, M.M., et al., 2013. Redes neuronales artificiales. Predicción de alturas del río paraná. FPUNE Sci. 8 (8).
Flores-López, F., Galaitsi, S., Escobar, M., Purkey, D., 2016. Modeling of Andean páramo ecosystems hydrological response to environmental change. Water 8 (3), 94.
Hermida, M., Hermida, C., Cabrera, N., Calle, C., 2015. La densidad urbana como variable de análisis de la ciudad: El caso de Cuenca, Ecuador. EURE (Santiago) 41 (124), 25–44.
Hernández, D.C., Ramírez, G.D., González, M.R., Caciano, R.T., Avalos, J.E., 2013. Ajuste y validación del modelo hidrológico GR2M en la cuenca alta del río nazas. Agrofaz 13 (2).
Hosseini, S.M., Mahjouri, N., 2016. Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling. Appl. Soft Comput. 38, 329–345.
Huard, D., Mailhot, A., 2008. Calibration of hydrological model GR2M using Bayesian uncertainty analysis. Water Resour. Res. 44 (2).
Iliadis, L.S., Maris, F., 2007. An Artificial Neural Network model for mountainous water-resources management: the case of Cyprus mountainous watersheds. Environ. Model. Softw. 22 (7), 1066–1072.
Jiang, P., Yu, Z., Gautam, M.R., Acharya, K., 2016. The spatiotemporal characteristics of extreme precipitation events in the western United States. Water Resour. Manage. 30 (13), 4807–4821.
Josse, C., Cuesta, F., Navarro, G., Barrena, V., Becerra, M.T., Cabrera, E., Chacón-Moreno, E., Ferreira, W., Peralvo, M., Saito, J., et al., 2011. Physical geography and ecosystems in the tropical Ande. Herzog, S.K., Martínez, R., Jørgensen, P.M., Tiessen, H. (comps.), Climate Change and Biodiversity in the Tropical Andes. Instituto Interamericano para la Investigación del Cambio Global y Comité Científico sobre Problemas del Medio Ambiente, São José dos Campos y París.
Kendall, M., 1975. Rank Correlation Methods, fourth ed. Charles Griffin, San Francisco, CA, pp. 8.
Lasdon, L.S., Waren, A.D., Jain, A., Ratner, M., 1976. Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming, Technical Report. Stanford Univ CA Systems Optimization Lab.
Lavado Casimiro, W.S., Labat, D., Guyot, J.L., Ardoin-Bardin, S., 2011. Assessment of climate change impacts on the hydrology of the Peruvian Amazon-Andes Basin. Hydrol. Process. 25 (24), 3721–3734.
Lema Changoluisa, M.A., Plaza Quezada, V.C., 2009. Modelación hidrológica de la cuenca (alta y media) del río pastaza aplicando el modelo de simulación WEAP (Water Evaluation and Planning System) (B.S. thesis). QUITO/EPN/2009.
Li, Z., Yu, J., Xu, X., Sun, W., Pang, B., Yue, J., 2018. Multi-model ensemble hydrological simulation using a BP neural network for the Upper Yalongjiang River Basin, China. Proc. Int. Assoc. Hydrol. Sci. 379, 335–341.
Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Model. Softw. 15 (1), 101–124.
Makhlouf, Z., Michel, C., 1994. A two-parameter monthly water balance model for French watersheds. J. Hydrol. 162 (3–4), 299–318.
Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 245–259.
McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash-Sutcliffe efficiency index. J. Hydrol. Eng. 11 (6), 597–602.
McLeod, A.I., 2005. Kendall Rank Correlation and Mann–Kendall Trend Test. R Package Kendall.
Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.
Mouelhi, S., 2003. Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. Université Paris VI, Paris, France.
Mouelhi, S., Michel, C., Perrin, C., Andréassian, V., 2006. Stepwise development of a two-parameter monthly water balance model. J. Hydrol. 318 (1–4), 200–214.
Mounir, Z.M., Ma, C.M., Amadou, I., 2011. Application of water evaluation and planning (WEAP): a model to assess future water demands in the Niger River (in Niger Republic). Mod. Appl. Sci. 5 (1), 38.
Mu noz, P., Orellana-Alvear, J., Willems, P., Célleri, R., 2018. Flash-flood forecasting in an Andean mountain catchment – development of a step-wise methodology based on the random forest algorithm. Water 10 (11), 1519.
Perea, R.G., Poyato, E.C., Barrios, M.P.M., Díaz, J.R., 2016. Optimización de la predicción de demanda de agua mediante algoritmos neuro-genéticos para un conjunto de datos reducido. XXXIV Congreso Nacional de Riegos 1.
Pereira, A.R., Pruitt, W.O., 2004. Adaptation of the Thornthwaite scheme for estimating daily reference evapotranspiration. Agric. Water Manage. 66 (3), 251–257.
Pourrut, P., Gómez, G., 1998. El ecuador al cruce de varias influencias climáticas. Una situación estratégica para el estudio del fenómeno El Niño. Bulletin de l'Institut français d'études andines 27 (3).
Ritter, A., Muñoz-Carpena, R., 2013. Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments. J. Hydrol. 480, 33–45.
Sarmiento, G., 1986. Ecological features of climate in high tropical mountains. High Alt. Trop. Biogeogr. 11, 45.
Si-min, Q., Wei-min, B., Peng, S., Zhongbo, Y., Peng, J., 2009. Water-stage forecasting in a multitributary tidal river using a bidirectional Muskingum method. J. Hydrol. Eng. 14 (12), 1299–1308.
Singh, M., 2018. Development of Integrated Water Resources Management Plan for a Basin Using Water Evaluation and Planning (Weap) Model (PhD thesis). IIT, Kharagpur.

Thornthwaite, C., et al., 1954. A re-examination of the concept and measurement of potential evapotranspiration. Publ. Climatol. 7 (1), 200–209.
Tkacz, G., Hu, S., et al., 1999. Forecasting GDP Growth Using Artificial Neural Networks. Bank of Canada, Ottawa.
Veintimilla-Reyes, J., Cisneros, F., Vanegas, P., 2016. Artificial neural networks applied to flow prediction: a use case for the Tomebamba river. Procedia Eng. 162, 153–161.
Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution metropolis algorithm for estimating posterior distribution of watershed model parameters. Calibration Watershed Models 105–112.
Wang, J., Shi, P., Jiang, P., Hu, J., Qu, S., Chen, X., Chen, Y., Dai, Y., Xiao, Z., 2017. Application of BP neural network algorithm in traditional hydrological model for flood forecasting. Water 9 (1), 48.
Weimin, B., Wei, S., Simin, Q., 2013. Flow updating in real-time flood forecasting based on runoff correction by a dynamic system response curve. J. Hydrol. Eng. 19 (4), 747–756.
Yates, D., Sieber, J., Purkey, D., Huber-Lee, A., 2005. WEAP21 – a demand-, priority-, and preference-driven water planning model: part 1: model characteristics. Water Int. 30 (4), 487–500.
Yi, S., Sandoval Solis, S., 2015. Hydrological WEAP modeling of the Upper Basin of the Apurimac River Basin, in Peru. An assessment of available water resources and supply. AGU Fall Meeting Abstracts.
Ytoui, Y., 2014. Rainfall-runoff modeling at monthly and daily scales using conceptual models and neuro-fuzzy inference system. 2nd International Conference – Water Resources and Wetlands 263–270.
Zhong, T., Xie, T., 2012. Application and simulation of matlab neural network tool NNTool. Comput. Mod 12, 44–47.
Zolfaghari, M., Mahdavi, M., Rezaei, A., Salajegheh, A., 2013. Evaluating GR2M model in some small watersheds of Iran (Case study Gilan and Mazandaran Provinces). J. Basic Appl. Sci. Res. 3 (2), 463–472.