Assessing Combinations of Artificial Neural Networks Input Meteorological Parameters to Improve Daily Runoff Simulation.

Yamina Aoulmi[1*], Nadir Marouf[2] and Mohamed Amireche[3]

[1*,2,3] Department of hydraulics, Faculty of technlogy
Larbi-Ben-M'hidi university Oum-El-Bouaghi, Algeria
[1]yaminaaoulmi453@gmail.com
[2]marouf_n@yahoo.fr
[3]amiremoh@hotmail.com

Abstract. Hydrological models are one of the tools used for reconstruction or simulation, forecasting for the anticipation of future changes in the flow of a river, which allows better management of water resources during low flow periods and anticipation of flood risks during high water periods. Using artificial neural network (ANN) in the fields of hydrology and water resources, became advocated due to its capability of tackling, modeling and forecasting the problems that are nonlinear or stochastic within the Rainfall-Runoff (R-R) system. This research aims to test the practicability of using ANNs with two input configurations; to model the R-R relationship in two stations Mirebek and Ain Berda located in the Seybouse basin in Algeria. The 1st (ANN I_P) considers only precipitation as an input variable for the daily flow simulation. The 2nd (ANN $II_{P,T,Hu}$) considers a combination of the Temperature and Humidity with precipitation at the model input. The results of the two models were compared through performance metrics, viz., Root Mean Square Error (RMSE),mean absolute error (MAE), Pearson's correlation coefficient (R), Nash Sutcliffe Efficiency coefficient (NSE), and through graphical interpretation (scatter plots and time series). Better flow simulations were provided by the three-input model for the two stations; where R=0.90, NSE=80.5% for Mirebek station and R=0.85, NSE=72% for Ain Berda. This result has confirmed that as much input variables are numerous, as more the model of ANN is efficient. The finding of this study indicates that the developed ANN models could be considered as a powerful tool for predicting Runoff.

Keywords: ANN, Predicting runoff, Seybouse basin, Meteorological parameters.

1 Introduction

The Seybouse basin with an area of 6,471 km^2 is one of five most extensive and largest basins in Algeria [1]. However, the increase in water needs linked to population growth, urbanization development, industrial and agricultural needs, also, the inefficient and improper use of available water resources, will cause afterward an increase in water demand compared to available water supply. Therefore, water control becomes a major preoccupation. An analysis of the water demand in long term for water resources management in the basin's 'Moyenne Seybouse' zone shows that, even if all the mobilizable resources will be used, the deficit will be felt more and more [2].

The modeling of the rainfall-runoff relationship is a very important research topic that can be used for the planning, operation and control of any water resources project. In most arid regions, especially where there is a lack of long-term hydrometeorological data, modeling has become a suitable tool for the management of water resources, both in wet and dry periods. In addition, future predictions with different scenarios in terms of climate change as well as the environment can be undertaken by rainfall-runoff modeling [3]. The rainfall-runoff modeling is highly nonlinear, stochastic and extremely complex process due to the basin characteristics and the various subprocesses that influence this complicated system. There are different models to forecast this process ,these models are classed as empirical / black box and conceptual or physical distributed models. Since distributed models based on physics are too complex, intensive, and bulky data to use, black box and conceptual models are commonly used for modelling rainfall-runoff [4]. Furthermore, the severe lack of environmental data such as geomorphological characteristics of the basin (topography, soil types, etc.) can prevent the use of conceptual models, which necessitate a large amount of data for calibration and validation purposes. Therefore, hydrological modelling of data scarce catchments become limited [5]. Thus, conceptual models for the rainfall runoff system do not apply to basins for which precise data and parameters are not available [6]. In such catchment, where limited data (e.g., rainfall and streamflow) are available, empirical models are highly efficient. As a result, we decided to concentrate on the black box or empirical models, which we stated earlier. The characteristic of this type of model is that it uses a mathematical function to establish a stable relationship between input and output variables without the use of complicated physical laws that control the natural mechanism.

Recently, a lot of attention has been focused on artificial intelligence-based (AI-based) data driven models [7], such as the Artificial Neural Network (ANN). The ANN technique is becoming extremely prevalent in the fields of hydrology and water resources, especially with the excellent results found in many studies; that demonstrated ANN ability and potential to model the intricate rainfall-runoff mechanism and to perform better than the conventional modelling techniques [8]–[12]. Some studies have shown the possibility of using other ANN input variables than precipitation to improve the established ANN's runoff prediction accuracy [13]**.**

As a hypothesis in this paper, it is speculated that simple changes to the ANNs input data will improve their output on flow simulation. It is well established, for example, that the flow is affected not only by total rainfall but also by other meteorological factors. As a result, we believe that using humidity and temperature as input parameters for ANNs can improve their efficiency.As a result of the foregoing, the aim of this article is to investigate the applicability of using ANNs with two input configurations for flow simulation in two station located in semi-arid zone (Seybouse basin). All the ANNs models were developed using daily precipitation, temperature, humidity, and flow data. To validate all the investigated models, a number of traditional statistical performance assessment methods as well as graphical analysis were used.

2 Methodology

2.1 Databases

For the two stations in the Seybouse basin, the collected database reflects regular sets of rainfall-runoff values, humidity, and temperature.

These data were taken from the databases of Algiers' National Agency of Water Resources (NAWR) and Constantine's National Office of Meteorology (NOM). Table 1 displays the two streams gauge stations located in the Seybouse basin.

Table 1. Hydrometric stations of Seybouse basin considered in this research.

Code	Name	River	Area (km^2)	Available Data period
14-06-01	Mirebek	Seybouse	5950	From 1981 to 1995
14-06-02	Ain Berda	Ressoul	102	From 1981 to 1997

2.2 Artificial Neural Network model training

The ANN (Fig. 1) structure developed in this research was the “MLP”, which is made-up of three layers, i.e. an input layer, a hidden (intermediate) layer and an output layer, this structure is formed with a back-propagation algorithm. Each layer contain different numbers of nodes that link to the nodes in the next layer with the connections called weights.

Input layer neurons transmit the input values to unprocessed hidden layer nodes. The values are distributed to all intermediate layer nodes according to the connection weights (Wij,Wjk) between the input node and the hidden nodes. Each node j receives incoming signals from each node i in the precedent layer. Each inlet signal (Xi) is associated with a Weight (Wij). The effective inlet signal (Sj) to node j is the weighted sum of all the incoming signals passing through an activation function, and bj is the neuron threshold value (Eq.1) [8]. The most commonly activation function used in this kind of network to generate the outgoing signal (y) from the node, is the tangent sigmoid (Eq. 2) which performance is better than the logistic sigmoid according to the results of [14] results.

$$Sj=\sum_{i=1}^{n} X_i W_{ij}+b_j \tag{1}$$

$$f(Sj)=\frac{2}{(1+e^{-2Sj})}-1 \tag{2}$$

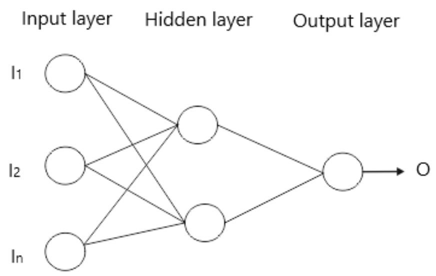

Fig. 1. The ANN Model Architecture

Two different configurations were considered when developing the ANNs models for the two gauging stations described earlier. The first was a traditional development (single entry model), which is an ANN that estimates daily flow from a single valid input variable that is solely related to the totals precipitation (ANN I_P). The second was an ANN that estimates daily flow using three input variables related to precipitation, temperature, and humidity (ANN $III_{P,T,Hu}$). The database was divided into three samples before these models were created: 70% for training, 15% for cross-validation, and 15% for testing. After several iterations of training the ANN to find the network with the smallest error and the highest performance criterion values, a final test is created to validate the network's selection. It entails calculating the Mean Absolute Error "MAE" by applying the chosen network to another database from the same station (which was not used in network training).

The simulation of the models built in this study was done using MATLAB (R2018b).

2.3 Performance evaluation indices

The following statistical criteria were adopted to assess the applied models. Theses statistical indices include root mean square error (RMSE); mean absolute error (MAE); Nash-Sutcliffe efficiency (NSE); and Pearson correlation coefficient (R). They are given by:

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(E_i-O_i)^2}{n}} \tag{3}$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n} \mid E_i - O_i \mid \tag{4}$$

$$NSE=\left[1-\frac{\sum_1^n (E_i-O_i)^2}{\frac{1}{n}\sum_1^n (O_i-\hat{O})^2}\right] \tag{5}$$

$$R = \sqrt{R^2} \quad \text{and} \quad R^2=1-\frac{\sum_1^n (O_i-E_i)^2}{\sum_1^n (O_i-\hat{O})^2} \tag{6}$$

Where: Ei and Oi are respectively, the estimated and observed value of flow, Ô is the mean of Oi values and n is the total data sets number.

3 Study zone

The Seybouse basin located in the northeast of Algeria (Fig. 2), has an area of 6471 km2. The main river, Oued-Seybouse with a total length of 240 km drains the entire surface of the basin, and constitutes an important source of water, used for all the needs of the region. The Seybouse basin is divided into 03 parts, namely: Haute-Seybouse, Moyen-Seybouse and Basse-Seybouse. The two study stations are distributed in the 'Basse-Seybousse' part.

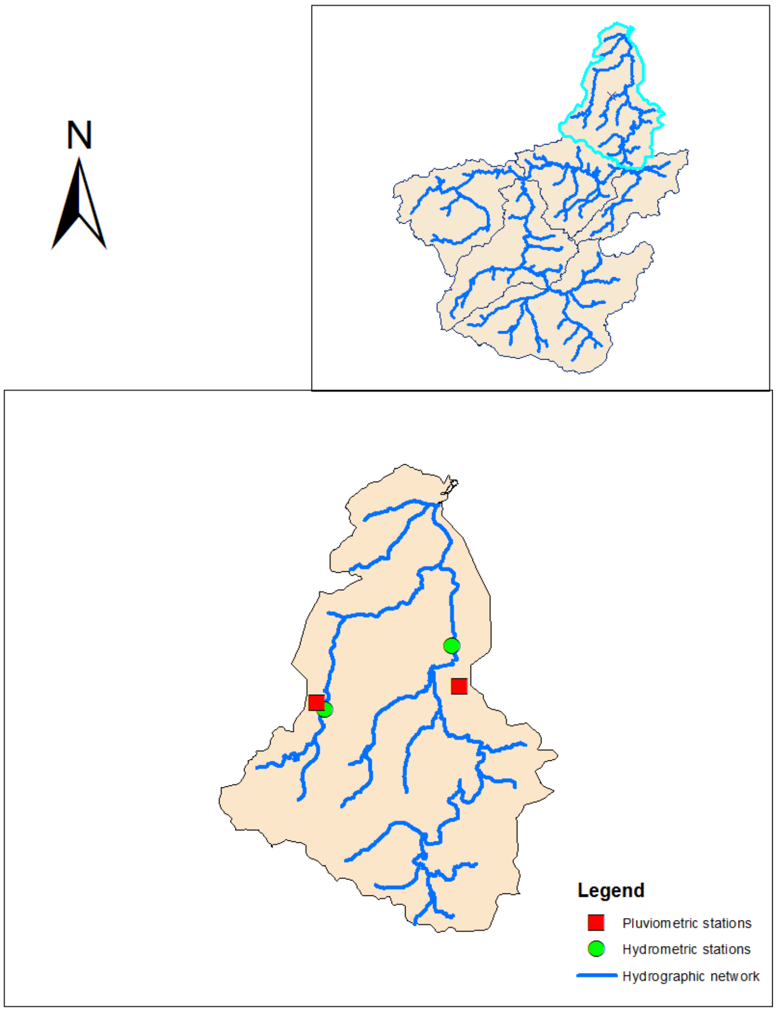

Fig. 2. The case study of the Seybouse basin

4 Results and discussion

In this chapter, the results in terms of differents performance statistics for the developed models is presented in the tables below.

The first model, which used a classical development to estimate regular flow, had only one input (the precipitation database); by varying the number of nodes in the middle layer, different ANN scenarios were developed. The final ANN scenario was chosen based on the highest values of performance parameters which found for a number of neurons in the intermediate layer equal to 30.

The values of the ANN I_P model performance indicators for the two study stations are shown in Table 3. According to this table, it can be seen that: the results obtained for Mirebeck station are not satisfactory; a Nash criterion equal to 74.7% a Pearson correlation coefficient varying from 0,71 to 0,87 in training, validation and testing phases of ANN I_P model, and a MAE whis is calculated between the two remaining samples of (P and Q) is equal to 0.14. Even Ain Berda station, the results wasn't acceptable; ; a NSE equal to 59.9% a R varying from 0,73 to 0,87 in training, validation and testing phases, and a MAE equal to 0.61.

Table 2. Statistical indices of the trained single input model (ANN I_P) applied to Mirebeck and Ain Berda stations.

Single input model	Performance criteria		Gauging stations	
			Mirebek	Ain Berda
ANN I_P	R	Training	0.8786	0.7326
		Validation	0.8592	0.8723
		Test	0.7189	0.8623
		All	0.8653	0.7737
	NASH		74.7	59.9
	RMSE		0.165	1.049
	MAE		0.1492	0.6141

Since the single-input ANN I_P models provided unsatisfactory results at both stations, a combination of two meteorological parameters (temperature and humidity) with precipitation seems necessary (ANN $III_{P,\ T,\ Hu}$), the results showed an improvement in performance criteria; where for Mirebeck station (Nash = 80.5%, R = 0.897, RMSE=0.14 and MAE=0.12) and for Ain Berda station (Nash = 72%, R = 0.85, RMSE=0.87 and MAE=0.55), the results are presented in Table 3.

Table 3. Statistical indices of the trained three input model (ANN $III_{P,T,Hu}$) applied to Mirebeck and Ain Berda stations.

Three input model	Performance criteria	Gauging stations	
		Mirebek	Ain Berda

ANN $III_{P,T,Hu}$	R	Training	0.9175	0.8241
		Validation	0.7277	0.9107
		Test	0.8036	0.9117
		All	0.897	0.8516
	NASH		80.5	72
	RMSE		0.144	0.877
	MAE		0.1233	0.5579

Also the graphical performance indicators present better results; where Fig. 3 shows he scatter plot alignment approaching the linear line y = x at 45 °, and Fig. 4 shows a very good overlap between the estimated and observed flows values in the training, validation and testing phases.

The results of applying the three-input model found for these two stations confirm the study made by [15] where he has shown that, it is always stated that the more input data of neural networks there are, the more the predictions approach the real values and the more efficient the models are.

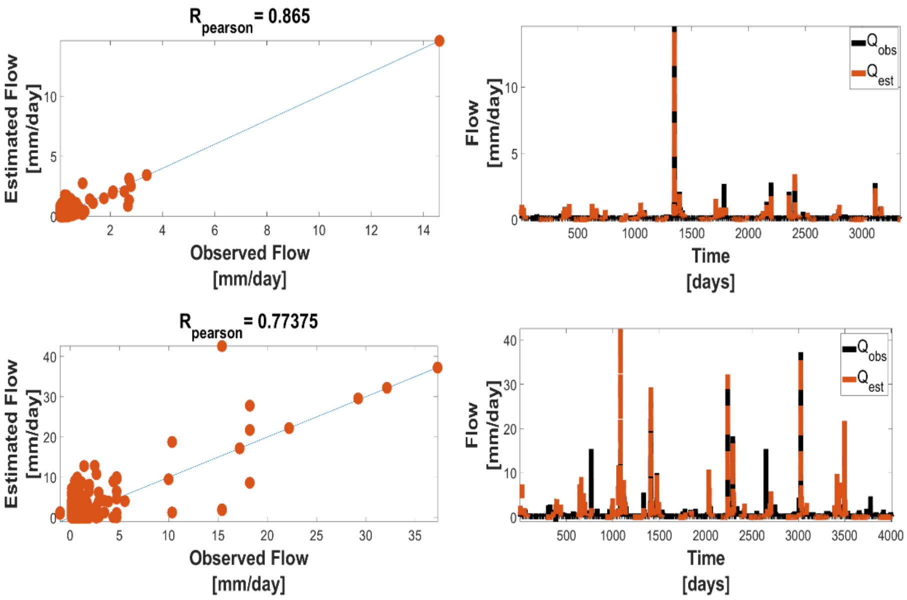

Fig. 7. Scatter plots and Time series of the observed and estimated daily flow values produced by ANN I_P for Mirebeck and Ain-Berda' stations.

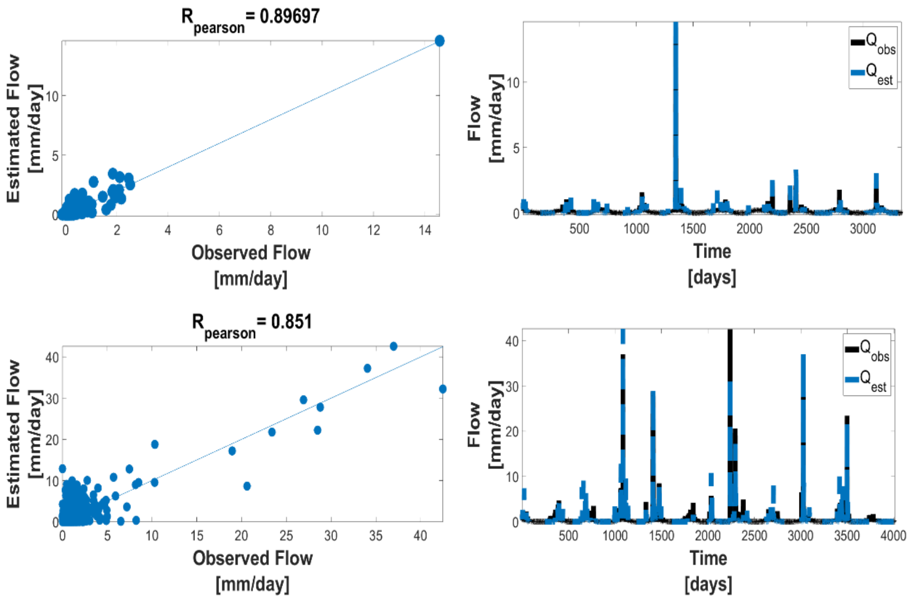

Fig. 8. Scatter plots and Time series of the observed and estimated daily flow values produced by ANN $III_{P,T,Hu}$ for Mirebeck and Ain-Berda' stations.

5 conclusion

This study addressed the complex process of rainfall-runoff modelling as well as real-time flow prediction using various configurations techniques.

Indeed, two main models were created for the two gauging stations spread within the 'Basse Seybouse' part, these models were optimized and compared to each other.The single-input models ANN I_P gave unsatisfactory results at the study stations. While the three-input model ANN $III_{P,T,Hu}$ provided a very good results compared to those obtained by previous models.

In semi-arid areas, ANN models have demonstrated their strength and capacity to simulate relatively accurate flows. The findings of this research showed that the more input variables there are, the more effective the ANN model is.

References

[1] ABH, Décret exécutif N°96-100 du 06 mars 1996 portant définition du bassin hydrographiques et fixant le status- type des établissements publics de gestion. .

[2] N. Brahmia et S. Chaab, « GESTION DES RESSOURCES EN EAU DANS LE BASSIN VERSANT DE LA MOYENNE SEYBOUSE », p. 4, 2013.

[3] F. S. Alahmadi, « RAINFALL-RUNOFF MODELING IN ARID REGIONS USING GEOGRAPHIC INFORMATION SYSTEMS AND REMOTE SENSING:CASE STUDY; WESTERN REGION OF SAUDI ARABIA », 2005. /paper/RAINFALL-

RUNOFF-MODELING-IN-ARID-REGIONS-USING-AND-Alahmadi/ee645f08d049dbe565ab99e792d459f74aedc73f (consulté le avr. 10, 2021).

[4] M. P. Rajurkar, U. C. Kothyari, et U. C. Chaube, « Artificial neural networks for daily rainfall—runoff modelling », Hydrological Sciences Journal, vol. 47, n° 6, p. 865‑877, déc. 2002, doi: 10.1080/02626660209492996.

[5] M. Jahandideh-Tehrani, G. Jenkins, et F. Helfer, « A comparison of particle swarm optimization and genetic algorithm for daily rainfall-runoff modelling: a case study for Southeast Queensland, Australia », Optim Eng, vol. 22, n° 1, p. 29‑50, mars 2021, doi: 10.1007/s11081-020-09538-3.

[6] R. S. Vilanova, S. S. Zanetti, et R. A. Cecílio, « Assessing combinations of artificial neural networks input/output parameters to better simulate daily streamflow: Case of Brazilian Atlantic Rainforest watersheds », Computers and Electronics in Agriculture, vol. 167, p. 105080, déc. 2019, doi: 10.1016/j.compag.2019.105080.

[7] V. Nourani, A. Hosseini Baghanam, J. Adamowski, et O. Kisi, « Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review », Journal of Hydrology, vol. 514, p. 358‑377, juin 2014, doi: 10.1016/j.jhydrol.2014.03.057.

[8] I. Aichouri, A. Hani, N. Bougherira, L. Djabri, H. Chaffai, et S. Lallahem, « River Flow Model Using Artificial Neural Networks », Energy Procedia, vol. 74, p. 1007‑1014, août 2015, doi: 10.1016/j.egypro.2015.07.832.

[9] K. Benzineb et M. Remaoun, « Daily rainfall-runoff modelling by neural networks in semi-arid zone: case of Wadi Ouahrane’s basin », Journal of Fundamental and Applied Sciences, vol. 8, n° 3, Art. n° 3, 2016, doi: 10.4314/jfas.v8i3.

[10] F. Machado, M. Mine, E. Kaviski, et H. Fill, « Monthly rainfall–runoff modelling using artificial neural networks », Hydrological Sciences Journal, vol. 56, n° 3, p. 349‑361, avr. 2011, doi: 10.1080/02626667.2011.559949.

[11] S. Srinivasulu et A. Jain, « A comparative analysis of training methods for artificial neural network rainfall–runoff models », Applied Soft Computing, vol. 6, n° 3, p. 295‑306, mars 2006, doi: 10.1016/j.asoc.2005.02.002.

[12] Z. M. Yaseen, « Artificial intelligence based models for stream-flow forecasting: 2000–2015 », Journal of Hydrology, p. 16, 2015.

[13] G.-F. Lin et G.-R. Chen, « A systematic approach to the input determination for neural network rainfall–runoff models », Hydrological Processes, vol. 22, n° 14, p. 2524‑2530, 2008, doi: 10.1002/hyp.6849.

[14] M. Rezaeianzadeh et al., « Assessment of a conceptual hydrological model and artificial neural networks for daily outflows forecasting », Int. J. Environ. Sci. Technol., vol. 10, n° 6, p. 1181‑1192, nov. 2013, doi: 10.1007/s13762-013-0209-0.

[15] R. N. Randrianarivony, P. Lauret, Z. A. Randriamanantany, et J. C. Gatina, « Modélisation du régime annuel des petites rivières en vue d’installation de microcentrales hydroélectriques », Afrique Science: Revue Internationale des Sciences et Technologie, vol. 5, n° 1, Art. n° 1, 2009, doi: 10.4314/afsci.v5i1.61705.