www.czasopisma.pan.pl PAN www.journals.pan.pl

JOURNAL OF WATER AND LAND DEVELOPMENT

e-ISSN 2083-4535

Polish Academy of Sciences (PAN) Institute of Technology and Life Sciences – National Research Institute (ITP – PIB)

JOURNAL OF WATER AND LAND DEVELOPMENT
DOI: 10.24425/jwld.2021.138158
2021, No. 50 (VI–IX): 38-47

The assessment of artificial neural network rainfall-runoff models under different input meteorological parameters Case study: Seybouse basin, Northeast Algeria

Yamina Aoulmi, Nadir Marouf, Mohamed Amireche

University of Larbi-Ben-M'hidi, Faculty of Sciences and Applied Sciences, Department of Hydraulic, Laboratory of Ecology and Environment, PO Box 358, 04000 Oum El Bouaghi, Algeria

RECEIVED 25.08.2020 REVIEWED 17.10.2020 ACCEPTED 03.11.2020

Abstract: Over the past two decades, artificial neural networks (ANN) have exhibited a significant progress in predicting and modeling non-linear hydrological applications, such as the rainfall-runoff process which can provide useful contribution to water resources planning and management. This research aims to test the practicability of using ANNs with various input configurations to model the rainfall-runoff relationship in the Seybouse basin located in a semi-arid region in Algeria. Initially, the ANNs were developed for six sub-basins, and then for the complete watershed, considering four different input configurations. The 1st (ANN I_P) considers only precipitation as an input variable for the daily flow simulation. The 2nd (ANN II) considers the 2nd variable in the model input with precipitation; it is one of the meteorological parameters (evapotranspiration, temperature, humidity, or wind speed). The third (ANN $III_{P,T,HUM}$) considers a combination of temperature, humidity, and precipitation. The last (ANN $V_{P,ET,T,HUM,Vw}$) consists in collating different meteorological parameters with precipitation as an input variable. ANN models are made for the whole basin with the same configurations as specified above. Better flow simulations were provided by (ANN $II_{P,T}$) and (ANN $II_{P,Vw}$) for the two stations of Medjez-Amar II and Bordj-Sabath, respectively. However, the (ANN $V_{P,ET,T,HUM,Vw}$)'s application for the other stations and also for the entire basin reflects a strategy for the flow simulation and shows enhancement in the prediction accuracy over the other models studied. This has shown and confirmed that the more input variables, as more efficient the ANN model is.

Keywords: artificial neural networks (ANNs), meteorological parameters, rainfall-runoff, semi-arid region, Seybouse basin, various input configurations

INTRODUCTION

The management of water resources is a major political, economic and social issue that governments and institutions identify as a priority on the political agenda in the 21st century [Bied-Charreton *et al*. 2004]. Currently, the challenge for Algeria, and more precisely for the Seybouse basin, is how to better meet user needs (domestic water supply, industry, agriculture) with relatively limited water resources; especially considering the inefficient and improper use of available water resources and high increase in demand due to population growth and industrial and agricultural development. Consequently, water resources in the Seybouse basin are subject to strong pressures which seriously threaten their sustainability. Hence, the interest remains high as regards the development of integrated management approaches that consider the watershed as a relevant management unit [Brahmia, Chaab 2013]. Hydrological modeling of the Seybouse basin can be a fundamental tool to improve its management and allows water resource managers to draw up an effective and rational strategy. Additionally, it allows a hydrological risk assessment linked to flood/drought designed to avoid losses in public and private property, human lives and tangible goods, health and ecological risks. Such risk are likely to occur due to disasters. Several models have been developed to simulate runoff forecasting. Runoff is used in many water management applications, such as forecasting extreme floods and dry periods, power generation, designing of hydraulic structures, and irrigation [SRINIVASULU, JAIN 2006]. These rainfall-runoff forecast models are classified as an empirical/black box and conceptual or physical distributed models. The theoretical black box and conceptual models are generally used for modeling rainfall-runoff, since distributed models based on physics are too complex and require intensive and cumbersome data [RAJURKAR *et al.* 2002]. Moreover, the significant lack of environmental data, such as geomorphological characteristics of the basin (topography, vegetation, soil types …), may hamper the use of conceptual models. Such models require large volumes of data for calibration and validation leading to computational inefficiency. Thus, the rainfall runoff mechanism deterministic or conceptual models cannot be applied to basins for which detailed data and parameters are not available [VILANOVA *et al.* 2019]. As a result, their use has been viewed with some skepticism by researchers and has not become very popular [GRAYSON *et al.* 1992]. This led us to focus on a separate category of models, i.e. the black box or empirical models. The application of black box models based on the ANN technique has become increasingly popular in hydrology and water management due to their ability and potential to provide satisfactory modeling of the intricate rainfall-runoff mechanism under limited data availability [AICHOURI *et al.* 2015; BENZINEB, REMAOUN 2016; MACHADO *et al.* 2011; SRINIVASULU, JAIN 2006; YASEEN 2015]. BHADRA *et al.* [2010] and REZAEIANZADEH *et al.* [2013] have compared the performance of different conceptual and ANN techniques for modeling and reported that the ANN approach which does not have physically realistic components and parameters, outperformed conventional conceptual models.

A multi-layer perceptron MLP-feed-forward is the most commonly used ANN in engineering applications [VIDYARTHI *et al.* 2020]. COULIBALY [1999] has mentioned that approximately 90% of hydrological neural network applications use multilayer feed-forward neural networks trained by the back propagation (BP) algorithm. For an efficient back propagation training, a Levenberg–Marquardt numerical optimization technique [LEVENBERG 1944; MARQUARDT 1963] can be incorporated into the back propagation algorithm to expedite and improve the training and to reach optimal solutions to several problems [BHADRA *et al.* 2010]. As a result, a three-layered feed-forward neural network, trained with backpropagation (FFBP) using Levenberg–Marquardt (LM) algorithms, has been applied in this study.

RANDRIANARIVONY *et al.* [2009] have concluded that ANNs represent very useful tools for rainfall-runoff models to fill the gap in survey data. However, the more data we have, the more realistic predictions can be achieved. In addition, some papers have shown the possibility of using other ANN input variables than precipitation to increase the developed ANN performance to predict runoff [LIN, CHEN 2008]. As a postulate and novelty, the article purports that a simple adjustment to ANN input data can be made to ameliorate their performance in flow simulation. For example, the flow not only depends on total rainfall, but also on other meteorological parameters. Therefore, we believe that the use of evapotranspiration, humidity, temperature and wind velocity as input parameters for ANNs can have a positive impact on their performance. The major goal of the article is to investigate the application of ANNs with different input configurations to simulate flow in a semi-arid area (Seybouse basin case). The daily average precipitation, temperature, potential evapotranspiration, humidity, wind velocity and flow data derived from the Seybouse basin have been employed to develop all the ANNs models. A variety of standard statistical performance evaluation measures and graphical performance indicators have been employed to validate all the investigated models. The paper begins with a brief presentation of the study area followed by the description of material and methods where data, details of the model's development, and performance criteria are presented. Results are discussed in the next chapter before concluding remarks.

MATERIAL AND METHODS

STUDY AREA

The Seybouse River basin is located the North-East of Algeria (Fig. 1). With an area of 6,471 km^2, it is one of the constituent parts of Constantinois-Seybouse-Mellegue, a large hydrographic basin [Décret exécutif N°96-100]. The basin is one of the principal collectors of rainwater from extreme N-E Algerian regions. It extends south-west over a distance of 160 km to the Saharan Atlas, and it reaches a maximum width of 120 km in its section at Jebel-Ouahch. The main river, wadi-Seybouse, of the total length of 240 km drains the entire surface of the basin. It constitutes an important water source, used for agricultural plains irrigation. It originates from Heractas and Sellaoua high plains and ends in Annaba coastal plain to flow into the Mediterranean Sea. It is formed by the confluence of Cherf and Bouhamdane Rivers at Madjez Amar and receives two other tributaries of unequal importance: wadi Mellah and wadi Ressoul Rivers. The Seybouse basin is divided into three parts, namely Haute-Seybouse, Moyenne-Seybouse and Basse-Seybouse. It is spread over seven provinces in eastern Algeria, and it covers the entire Wilaya of Guelma, and partially the following Wilaya: Oum-El-Bouaghi; Constantine; Skikda; Souk-Ahras; Annaba and El Taref. The area includes 68 municipalities of which 30 are fully within the area examined. Water resources are vital to support the majority of economic activities in the region. The basin's climate varies from the typical Mediterranean to semi-arid. Average annual precipitation ranges from 700 mm to 400 mm, reaching from 90 to 120 mm monthly levels in December–January. Minimum temperatures are in December–January (less than 10°C) and maximum in July or August (between 25 and 30°C). The average annual evapotranspiration is around 1371 mm while the surface runoff represents 79 mm·y^{-1}.

DATABASES

The compiled database represents daily sets of rainfall-runoff values and meteorological variables (evapotranspiration, humidity, temperature and wind velocity) for the Seybouse basin at different periods for each hydrometric station. Due to the unavailability of data at these stations for the same periods, the data were studied at different periods. Data were collected from the database available in the National Agency of Water Resources and National Office of Meteorology in Algeria (Fr. Agence Nationale des Ressources hydrauliques – ANRH and Office National de Météorologie – ONM) and POWER Data Access Viewer v2.0.0 (URL: https://power.larc.nasa.gov/data-access-viewer/). Table 1 shows data for gauge stations of the Seybouse basin.

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

Code	Name	River	Area (km^2)	Available data period
14-02-02	Moulin Rochefort	Cherf	1 710	1981–1994
14-03-01	Medjez-Amar II	Moyenne Seybouse	1 105	1981–2002
14-03-02	Bordj-Sabath	Bouham-dane	304	1981–2015
14-05-01	Bouchegouf	Melah	550	1981–1995
14-06-01	Mirebek	Seybouse	5 950	1981–1995
14-06-02	Ain Berda	Ressoul	102	1981–1997

Source: own elaboration.

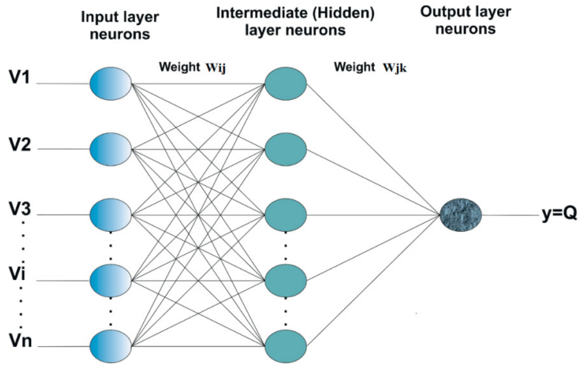

Fig. 2. Schematic representation of the artificial neural network (ANN) model architecture; V1, V2, V3, …, Vn = input variables used for the ANN (single, two, three, …, n) input models; i, j = nodes; y = outgoing signal; Q = flow; source: own elaboration

The six hydrometric stations cited above are distributed throughout the study basin, each one is provided with a rainfall gauge and with the same availability of data as the hydrometric station. The runoff/rainfall station classification is as follows: "Moulin Rochefort / Ain Makhlouf", "Medjez-Amar II / Medjez-Amar II", "Bordj-Sabath / Bordj-Sabath", "Bouchegouf / wadi Cheham", "Mirebek / Boukhamouza", "Ain Berda / Ain Berda".

Based on the ArcGis software, Figure 1 shows the Seybouse basin and its division into sub-basins, as well as the distribution of six hydrometric and rainfall stations.

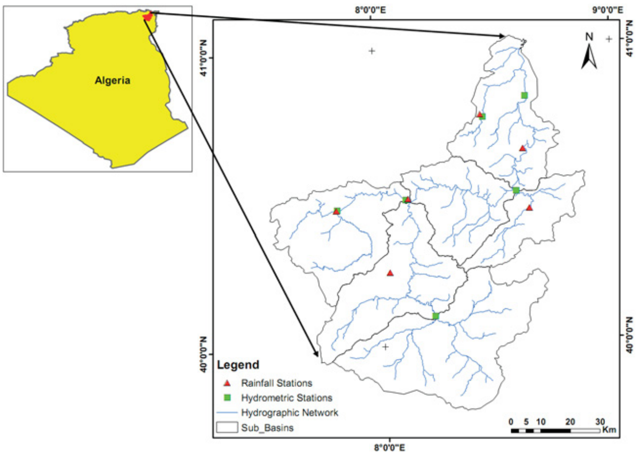

Fig. 1. Location of the Seybouse basin; source: own elaboration based on data of Hydrographic Basin Agency (Fr. Agence Nationale des Ressources hydrauliques – ABH)

ARTIFICIAL NEURAL NETWORK (ANN) TRAINING

Neural networks consist of simple elements (or neurons) that work in parallel. These elements are strongly inspired by the Nervous Biological System. An ANN (Fig. 2) carried-out by CorelDRAW X6 includes a set of components called artificial neurons, distributed among the layers that are mathematically interconnected via a transfer function [Machado *et al.* 2011]. The strength of the neuronal connection in adjacent layers is known as weight. There are different types of ANN, the most common of which is the ANN Multilayer Perceptron (MLP). the MLP is made-up of three layers, i.e. an input layer, a hidden (intermediate) layer and an output layer. Rezaeianzadeh *et al.* [2013] have applied the standard conceptual HEC-HMS's soil moisture accounting (SMA) algorithm and the multi-layer perceptron (MLP) for forecasting daily outflows at the Khosrow Shirin basin outlet, their findings indicate that the MLP can predict the daily flow with a higher forecasting efficiency than the conventional hydrological model HMS SMA. Additionally, the MLP is much simpler to apply than the HMS SMA by using a simple trial and error procedure. The ANN structure developed in this research was the "MLP" with a back-propagation algorithm to predict runoff from the drainage basin, while the neuron number in the middle layer was 30.

Input layer nodes that represent various input variables transmit input signals to unprocessed hidden layer nodes. These values are distributed to all intermediate layer nodes according to their connection weights (W_{ij}, W_{jk}) between the input node and hidden nodes. Each node j receives incoming signals from each node i in the precedent layer. Each inlet signal (X_i) is associated with a weight (W_{ij}). The effective inlet signal (S_j) to node j is the weighted sum of all the incoming signals passing through an activation function, and b_j is the neuron threshold value (Eq. 1) [Aichouri *et al.* 2015]. The activation function most commonly 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 Rezaeianzadeh *et al.* [2013].

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

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

ANN training use Levenberg–Marquardt algorithm, since it is more efficient, faster and has been highly recommended as the supervised algorithm of the first-choice. It is also used for identical purposes [Kashani *et al.* 2014].

All ANNs training and testing procedures were conducted using MATLAB software (R2018b).

The ANNs were developed for each sub-basin, where the above-mentioned gauging stations are spread in the study region.

Four different configurations were considered. The first consisted of a classical development (single input model), which is an ANN estimating the daily flow from a single relevant input variable exclusively related to the precipitated totals (ANN I_P). The second type was always an ANN estimating the daily flow, but its input variables included precipitation and one of the meteorological variables (model with two-inputs), consequently four sub-models with two-inputs were developed (ANN $II_{P,ET}$, ANN $II_{P,T}$, ANN $II_{P,HUM}$, ANN $II_{P,Vw}$). The third type was an ANN that uses three input variables related to precipitation, temperature and humidity to estimate the daily flow (ANN $III_{P,T,HUM}$). Finally, the last and fourth configuration consisted in developing an ANN with five input variables (precipitation, evapotranspiration, temperature, humidity and wind speed) to estimate the daily flow more accurately (ANN $V_{P,ET,T,HUM,\ Vw}$). All parameters used in various ANN model training are converted to the same measurement unit.

Before establishing these models, the database was split into three samples: 70% for training, 15% for cross-validation, and 15% for ANNs testing. After several series of ANN training to find the network that corresponds to the smallest error and to the largest performance values, the final test is developed to confirm the selection of the network. It involves another database from the same station (which is not used in network's training) and the calculation of the mean absolute error (*MAE*) by applying the chosen network (more details in section 4). Table 2 shows data sets from the different study stations used for ANN training, validation and testing, as well as the data used for the *MAE* test.

Table 2. Data subdivision for studied stations

Station name	Data period used for ANN training, validation and testing	Data period used for *MAE* test
Moulin Rochefort	1981–1985 and 1990–1994	1985–1990
Medjez-Amar II	1981–1988 and 1995–2002	1988–1995
Bordj-Sabath	1981–1992 and (2003–2015	1992–2003
Bouchegouf	1981–1985 and 1990–1995	1985–1990
Mirebek	1981–1985 and 1990–1995	1985–1990
Ain Berda	1981–1986 and 1992–1997	1986–1992

Source: own elaboration.

Any data used in a model to estimate the flow of one of the above-mentioned hydrometric stations had to be divided in the same manner as the simulated station (as shown in Tab. 2).

The modeling of different configurations is activated, once the data are entered and the network is created with its complete architecture (type, structure, function, layers, neurons). The models of different configurations described above, which are developed for all the stations regularly spread across the study area, are also developed for the entire basin.

MODEL PERFORMANCE EVALUATION INDICES

After developing the ANNs models, both statistical and graphic criteria were adopted to assess these models performance and select the most optimal desired model. Statistical indices included root mean square error (*RMSE*), mean absolute error (*MAE*), Nash–Sutcliffe efficiency (*NSE*), and Pearson correlation coefficient (*R*). They are given by:

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

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

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

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

where: E_i and O_i are respectively, the estimated and observed value of flow, is the mean of O_i values and n is the total data sets number.

The *RMSE* and *MAE* statistic measures are used to quantify the error between observed and estimated values. Pearson correlation coefficient *R* allows to measure the linear correlation between the observed and estimated flow values. The values range between −1 and +1; a negative value expresses the downhill linear relationship while a positive value expresses the uphill linear relationship.

The *NSE* criterion is a very popular index to assess the predictive power of a hydrological model. It can be interpreted as the observed flow variance proportion described by the model, ranging from −∞ to 100% [Nash, Sutcliffe 1970]. If *NSE* = 100% the model is perfect, if *NSE* = 90% the model is excellent, if it varies between 80% and 90% the model is very good, and if it ranges from 60% to 80% the model is good, under 60% the model is bad, but if *NSE* < 50, the flow calculated by the model is worse than the simple average flow [Chergui 2019].

The graphical performance indicator gives better results if the alignment of the scatter plot approaches the $y = x$ line at 45°, and desired (estimated) flow values overlap well with the observed flow values in both training and testing phases.

RESULTS AND DISCUSSION

To better appreciate performance and robustness of models developed, and therefore their predictive power, we present below the results in terms of various performance statistics of all ANNs models applied to the six studied stations. The first model, which consists of a classical development to estimate the daily flow, uses a single input (precipitation database as input). Different ANN scenarios have been formed by varying nodes in the middle layer. For each ANN scenario, the network has been trained. The training session was interrupted after each iteration to check the performance error (*MSE*) determined for training and testing phases. The training was finalized in the iteration that referred to the lowest network error calculated using the validation sample. The final ANN scenario was selected based on the maximum possible performance criteria that were found for a number of nodes in the intermediate layer equal to 30.

Once the network is selected, a last check confirms the chosen scenario performance. It consists in applying the Network found, to another precipitation sample P recorded from the same station to define a new variable y: y = network (P'), then calculating the mean absolute error (*MAE*) between this variable and the flow sample Q: *MAE* = mean2 (abs (y – Q')). This error should be minimal to confirm our selection and have a good network. The values of the most efficient ANN I_P performance indicators for each of the six stations studied are shown in Table 3. According to the table, it can be seen that results obtained for Medjez-Amar II station are very satisfactory; a *NSE* criterion is larger than 88% and a very strong Pearson correlation coefficient varies from 0.91 to 0.96 in training, validation and testing phases of ANN I_P. Even the *MAE* calculated between the two remaining samples of (*P* and *Q*) confirms the efficiency of this model with a minimal value of 0.74. However, this model (single input) results for the Ain Berda station are the worst compared with the other stations *NSE* < 60% and *R* = 0.77. The other stations results are satisfactory (a good *NSE* varies between 66 and 83 and a very strong *R* between 0.81 to 0.91).

The combination of another input parameter and precipitation improves the performance of precedent neural network models for the stations, no matter which input parameter is used (evapotranspiration *ET*, temperature *T*, humidity *HUM* or wind velocity *Vw*). For the Moulin Rochefort station, the use of humidity (*HUM*) as the 2nd input variable provided good results (*NSE* = 68.5% and *R* = 0.82) compared with models which used evapotranspiration (*ET*), temperature (*T*) or wind velocity (*Vw*), while for the Bordj-Sabath station, the use of wind speed gave better results (*NSE* = 88% and *R* = 0.93). The temperature influence was the most dominant in the remaining stations. Hence, the use of this meteorological parameter at these stations showed more influence on the daily flow oscillations. The Ain Berda station, which provided unsatisfactory results in the previous model, was improved with the ANN $II_{P,T}$ model with the *NSE* good (74%) and Pearson coefficient very strong (0.86). The improvement in the remaining stations was even better. Results are included in Table 4.

Since the two-input models of ANN $II_{P,T}$ and ANN $II_{P,HUM}$ provided better results compared with the other stations, the combination of temperature and humidity with precipitation was used to form a network with three-inputs (ANN $III_{P,T,HUM}$). Results showed improvements in the performance criteria for the stations of Moulin Rochefort (*NSE* = 70% and *R* = 0.83) and Bouchegouf (*NSE* = 84% and *R* = 0.91), while the remaining stations showed a deterioration in their performance compared with the two-input models. Results are presented in Table 5.

To optimize neural models for different stations, we decided to add precipitation and combine the parameter with four other meteorological parameters (*ET*, *T*, *HUM* and *Vw*) used in the preceding models. Thus, a network of five input variables was created. The ANN $V_{P,ET,T,HUM,Vw}$ model developed for each station was optimized and compared with the other models (single-input model ANN I_P, two-input model ANN II and three-input model ANN $III_{P,T,HUM}$). The results presented in Table 6 indicate a noticeable improvement in neural network performance. In fact, in most of the studied stations, we obtained a considerable increase in performance criteria compared with models discussed above (e.g. for Ain Berda station, the NSE criterion increased to 82% and Pearson coefficient to 0.9). Results obtained by this model were very satisfactory which proved its proper functioning.

RANDRIANARIVONY *et al.* [2009] have shown that the more neural network input data are available, predictions are closer to real values and models are more efficient.

A five-input model is not required to compare Tables 4 and 6 for Medjez-Amar II and Bordj-Sabath stations. Consequently, development of two models with two-inputs (ANN $II_{P,T}$ and ANN $II_{P,Vw}$) for the stations, respectively, is more than sufficient to achieve very good results.

Since the ANN $V_{P,ET,T,HUM,Vw}$ model has shown very good results in most stations regarding statistical criteria, we have graphically presented these results for each station to confirm their performance.

Figure 3 (A_I, B_I, C_I, D_I, E_I, F_I) shows a scatter plot alignment approaching the $y = x$ line at 45° for the all the gauging stations.

Table 3. Statistical indices of the trained single input model (ANN I_P) applied to the six hydrometric stations distributed in the Seybouse basin

Performance criteria		Gauging station					
		Moulin Rochefort	Medjez-Amar II	Bordj-Sabath	Bouchegouf	Mirebek	Ain Berda
$R_{Pearson}$	training	0.80	0.94	0.90	0.80	0.87	0.73
	validation	0.85	0.96	0.94	0.95	0.85	0.87
	test	0.78	0.91	0.93	0.93	0.71	0.86
	all	0.81	0.93	0.91	0.87	0.86	0.77
NSE		66.1	88.2	83.2	77.1	74.7	59.9
RMSE		0.05	0.40	0.52	0.51	0.16	1.04
MAE		0.04	0.74	0.59	0.30	0.14	0.61

Explanations: *R* = Pearson correlation coefficient, *NSE* = Nash–Sutcliffe efficiency, *RMSE* = root mean square error, *MAE* = mean absolute error.
Source: own study.

Table 4. Statistical indices of the trained two-input model (ANN $II_{P,ET}$, ANN $II_{P,T}$, ANN $II_{P,HUM}$, ANN $II_{P,Vw}$) applied to the six hydrometric stations distributed in the Seybouse basin

Performance criteria			Gauging stations					
			Moulin Rochefort	Medjez-Amar II	Bordj-Sabath	Bouchegouf	Mirebek	Ain Berda
ANN $II_{P,ET}$	$R_{Pearson}$	training	0.81	0.93	0.91	0.87	0.91	0.81
		validation	0.91	0.97	0.93	0.95	0.80	0.81
		test	0.77	0.96	0.96	0.90	0.72	0.90
		all	0.82	0.95	0.93	0.89	0.88	0.83
	NSE		67.4	90.7	86.5	80.5	78.9	69.2
	RMSE		0.04	0.35	0.46	0.47	0.15	0.91
	MAE		0.03	0.67	0.40	0.31	0.13	0.94
ANN $II_{P,T}$	$R_{Pearson}$	training	0.83	0.95	0.92	0.86	0.91	0.84
		validation	0.83	0.92	0.94	0.95	0.87	0.88
		test	0.72	0.97	0.93	0.90	0.81	0.90
		all	0.82	0.95	0.93	0.89	0.90	0.86
	NSE		67.9	91.5	86.6	80.7	81.5	74.1
	RMSE		0.04	0.34	0.46	0.47	0.14	0.84
	MAE		0.03	0.76	0.18	0.22	0.12	0.53
ANN $II_{P,HUM}$	$R_{Pearson}$	training	0.83	0.92	0.93	0.88	0.91	0.84
		validation	0.85	0.96	0.92	0.80	0.81	0.90
		test	0.71	0.98	0.90	0.92	0.74	0.90
		all	0.82	0.95	0.92	0.88	0.89	0.85
	NSE		68.5	91.0	86.0	77.6	80.6	73.6
	RMSE		0.04	0.35	0.47	0.51	0.14	0.85
	MAE		0.03	0.69	0.56	0.29	0.15	0.56
ANN $II_{P,Vw}$	$R_{Pearson}$	training	0.82	0.93	0.93	0.83	0.89	0.80
		validation	0.77	0.90	0.96	0.95	0.72	0.87
		test	0.84	0.98	0.90	0.92	0.76	0.84
		all	0.81	0.95	0.93	0.88	0.87	0.83
	NSE		67.0	90.6	88.0	78.7	76.7	69.8
	RMSE		0.04	0.35	0.43	0.50	0.15	0.91
	MAE		0.04	0.85	0.68	0.33	0.14	0.87

Explanations as in Tab. 3.
Source: own study.

Estimated and observed flows nearly overlapped during training, validation and testing for all the basin stations (Fig. 3 A_{II}, B_{II}, C_{II}, D_{II}, E_{II}, F_{II}).

These graphic performance indicators confirm the effectiveness of the ANN $V_{P,ET,T,HUM,Vw}$ model.

The flow modeling in six stations distributed throughout the basin, as well as the good results produced by the majority of the models, encouraged us to provide modeling for the whole basin. For this purpose, we gathered databases for all the stations to establish various models described above (model with single input, two-inputs, three-inputs and five-inputs). For the first model (single-input), the input variable was a grouping of precipitation data from the six stations used to estimate the daily flow for the entire basin. As explained above, before we started the ANN training, the set of data gathered (*P* and *Q* from the various stations) to form the entire sample was divided into three samples as follows: 70% for training, 15% for cross-validation, and 15% for ANNs testing. The other multiple input models were developed in the same manner, i.e. the meteorological database was used as input variables combined with precipitation, after collecting all data from the stations. Performance criteria for different models applied in the basin are presented in Table 7.

The model with a single input provided good results (*NSE* = 77% and *R* = 0.88).

The second input parameter improved results achieved (no matter which meteorological variable was used, including *ET*, *T*, *HUM* and *Vw*. All of them provided almost very close results, with *NSE* from 77 to 79% and *R* from 0.88 to 0.89).

Table 5. Statistical indices of the trained three-input model (ANN $III_{P,T,HUM}$) applied to the six hydrometric stations distributed in the Seybouse basin

Performance criteria		Gauging station					
		Moulin Rochefort	Medjez-Amar II	Bordj-Sabath	Bouchegouf	Mirebek	Ain Berda
$R_{Pearson}$	training	0.83	0.95	0.91	0.92	0.91	0.82
	validation	0.79	0.80	0.92	0.82	0.72	0.91
	test	0.86	0.83	0.89	0.94	0.80	0.91
	all	0.83	0.93	0.90	0.91	0.89	0.85
NSE		70.0	88.2	81.4	84.0	80.5	72.0
RMSE		0.04	0.40	0.54	0.43	0.14	0.87
MAE		0.04	0.75	0.51	0.29	0.12	0.55

Explanations as in Tab. 3.
Source: own study.

Table 6. Statistical indices of the trained five-input model (ANN $V_{P,ET,T,HUM,Vw}$) applied to six hydrometric stations distributed in the Seybouse basin

Performance criteria		Gauging station					
		Moulin Rochefort	Medjez-Amar II	Bordj-Sabath	Bouchegouf	Mirebek	Ain Berda
$R_{Pearson}$	training	0.84	0.93	0.92	0.93	0.93	0.90
	validation	0.83	0.97	0.81	0.88	0.88	0.88
	test	0.89	0.97	0.95	0.93	0.87	0.88
	all	0.85	0.95	0.92	0.93	0.92	0.90
NSE		72.9	90.8	85.0	86.6	85.1	82.0
RMSE		0.04	0.35	0.49	0.39	0.12	0.70
MAE		0.04	1.08	0.46	0.25	0.13	0.88

Explanations as in Tab. 3.
Source: own study.

Table 7. Statistical indices of the models applied for the entire Seybouse basin

Performance criteria		Values for ANN model						
		ANN I_P	ANN $II_{P,ET}$	ANN $II_{P,T}$	ANN $II_{P,HUM}$	ANN $II_{P,Vw}$	ANN $III_{P,T,HUM}$	ANN $V_{P,ET,T,HUM,Vw}$
$R_{Pearson}$	training	0.86	0.87	0.87	0.90	0.88	0.88	0.91
	validation	0.92	0.90	0.90	0.86	0.89	0.91	0.87
	test	0.90	0.89	0.90	0.84	0.88	0.85	0.90
	all	0.88	0.88	0.88	0.89	0.88	0.88	0.91
NSE		77.9	77.7	78.4	79.4	78.2	79.0	83.0
RMSE		0.69	0.69	0.68	0.66	0.68	0.67	0.60
MAE		0.49	0.47	0.46	0.47	0.46	0.46	0.49

Explanations as for Tab. 3.
Source: own study.

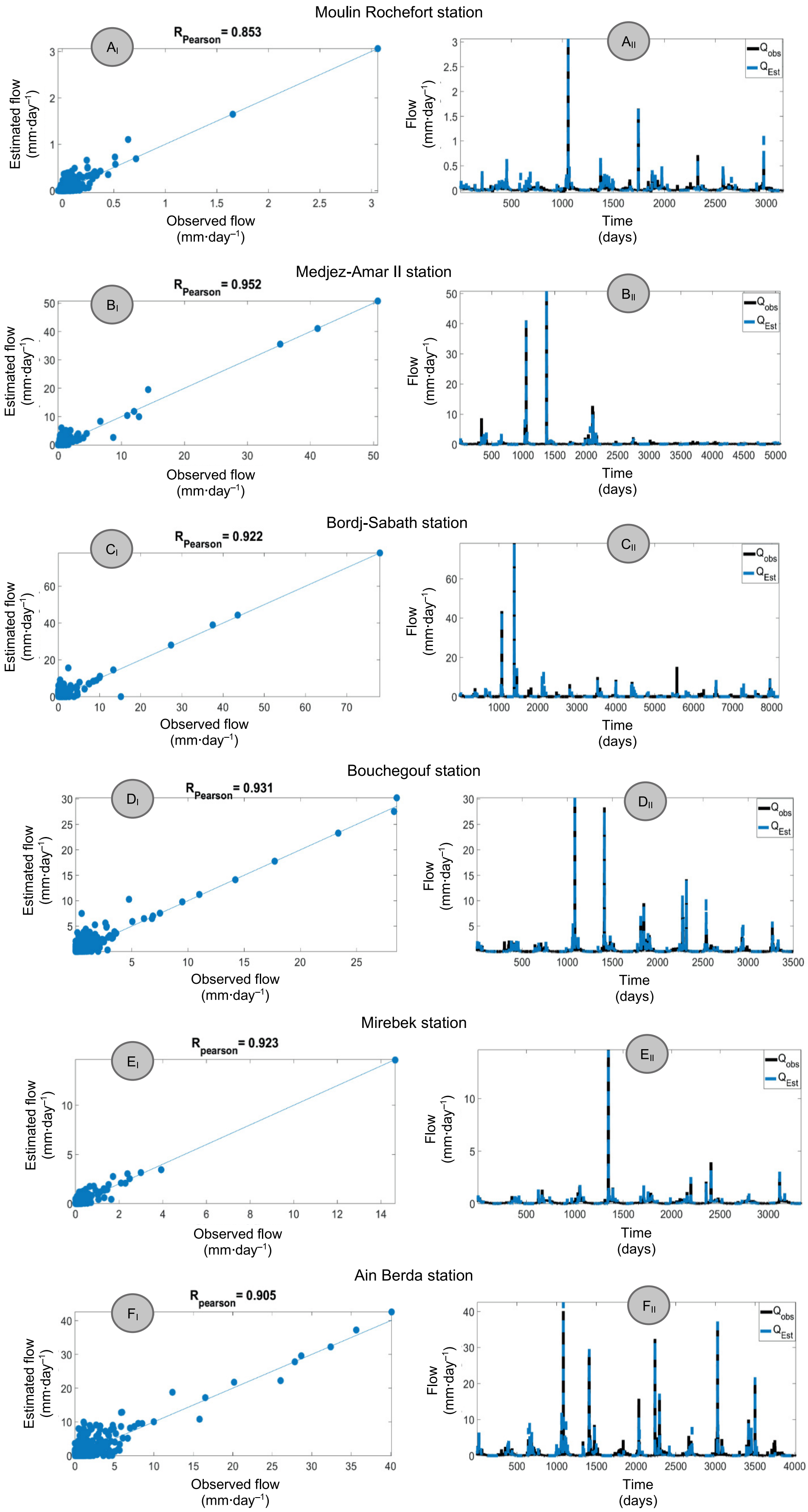

Fig. 3. Comparison of observed and estimated flow for the ANN V model in six gauging stations of the Seybouse basin; Q_{obs} = observed flow, Q_{est} = estimated flow; source: own study

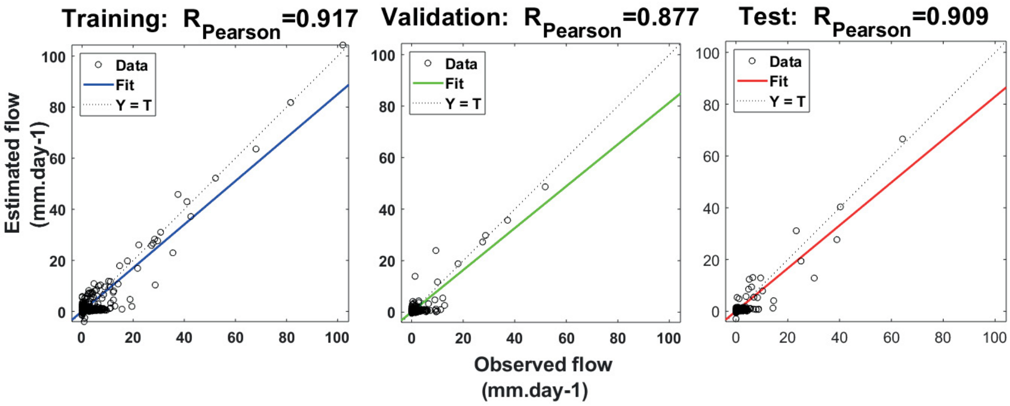

Fig. 4. Scatter plot for the model ANN $V_{P,ET,T,HUM,Vw}$ in training, validation and testing phases; source: own study

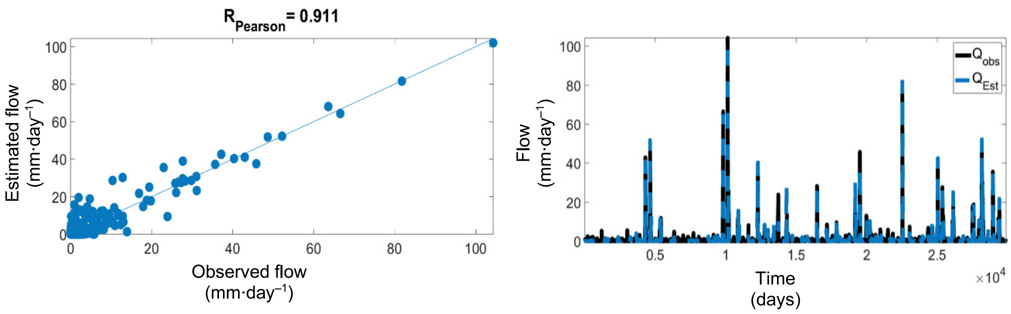

Fig. 5. Comparison of observed and estimated flow for the model ANN $V_{P,ET,T,HUM,Vw}$ of the Seybouse basin; source: own study

The three-input model (combination of temperature and humidity with precipitation) also improved results compared with the single-input model. However, its results were very close to that of the two-input model. We decided to improve the efficiency of neural network models and used the five-input model. The model generated very satisfactory results regarding static performance criteria (*NSE* = 83% and Pearson correlation coefficient 0.9) and graphic performance indicators (estimated and observed flow values nearly overlapped, and the alignment of the scatter plot approached the $y = x$ line at 45°) – Figures 4 and 5.

CONCLUSIONS

This article tests the practicability of artificial neural networks (ANNs) with different input configurations. The networks were used to simulate a rainfall-runoff model to better understand the hydrological behavior of the Seybouse basin. In fact, four main models were created for the six gauging stations spread within the study area. In order to estimate the daily flow, these models were optimized and compared with each other. The single-input model (precipitation) provided very satisfactory results at the Medjez-Amar II station, bad for Ain Berda, and acceptable results for the remaining stations.

The addition of a second meteorological parameter to the set in the previous ANN model improved its performance in all the stations studied. For the Moulin Rochefort station, the use of humidity as 2nd input variable provided good results compared with models that used the other meteorological parameters, while for the Bordj-Sabath station, the model that combined wind speed with precipitation as input variables provided better results. However, the influence of temperature in this type of a model was the most dominant for the rest of the stations and the model provided good results. To enhance the results, the combination of two meteorological parameters (humidity and temperature) with precipitation was used as input to a new model. Indeed, this neural model improved the results only in Moulin Rochefort and Bouchegouf stations, while the remaining stations showed deteriorated performance compared with the two-input model. To improve previous results and to optimize the ANNs models, the use of all meteorological variables and their combination with precipitation seems necessary. The model with five inputs developed for various stations provided much better results than those obtained by previous models. The last part of the study was to apply four previous configurations to the whole basin. Finally, the five-input model showed very satisfactory results.

ANN models have shown their power and capacity to simulate reasonably correct flows in semi-arid regions. The results and the comparative study of the different input configurations indicate that as much input variables are numerous, as more the model of ANN is efficient; therefore, the proposed ANN models are recommended for rainfall-runoff modeling, because of their simple structures, and their precision which helps us to solve problems related to water resources management.

REFERENCES

Aichouri I., Hani A., Bougherira N., Djabri L., Chaffai H., Lallahem S. 2015. River flow model using artificial neural networks. Energy Procedia. Vol. 74 p. 1007–1014. DOI 10.1016/j.egypro.2015.07.832.

BENZINEB K., REMAOUN M. 2016. 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 p. 956–970. DOI 10.4314/jfas.v8i3.17.

BHADRA A., BANDYOPADHYAY A., SINGH R., RAGHUWANSHI N.S. 2010. Rainfall-runoff modeling: comparison of two approaches with different data requirements. Water Resources Management. Vol. 24 p. 37–62. DOI 10.1007/s11269-009-9436-z.

BIED-CHARRETON M., PETIT O., MAKKAOUI R. 2004. La gouvernance des ressources en eau dans les pays en développement [Governance of water resources in developing countries]. Cahier du C3ED. No. 04-01. Centre d'Economie et d'Ethique pour l'Environnement et le Développement, Université de Versailles St-Quentin-en-Yvelines pp. 43.

BRAHMIA N., CHAAB S. 2013. Gestion des ressources en eau dans le bassin versant de la Moyenne Seybouse. En: Proceeding du Séminaire International sur l'Hydrogéologie et l'Environnement SIHE 2013 Ouargla [Management of water resources in the Moyenne Seybouse watershed. In: Proceeding of the International Seminar on Hydrogeology and Environment SIHE 2013 Ouargla] [online] p. 127–130. [Access 20.07.2020]. Available at: https://dspace.univ-ouargla.dz/jspui/bitstream/123456789/11953/1/35.pdf

CHERGUI A. 2019. Modelisation pluie-debit par reseaux de neurones artificiels du bassin versant de Sybousse [Rainfall-runoff modeling by artificial neuron networks of the Sybousse watershed]. MSc Thesis. Oum El-Bouaghi, Algeria. Université L'Arbi Ben M'hidi pp. 100 + Annexes.

COULIBALY P. 1999. Prévision hydrologique par réseaux de neurones artificiels : état de l'art [Hydrological forecasting by artificial neural networks: state of the art]. Canadian Journal of Civil Engineering. Vol. 26. Iss. 3 p. 293–304. DOI 10.1139/l98-069.

Décret exécutif N°96-100 du 17 Chaoual 1416 correspondant au 06 mars 1996 portant définition du bassin hydrographiques et fixant le status- type des établissements publics de gestion [Executive Decree No. 96-100 of 6 March 1996 defining the hydrographic basin and setting the standard status of public management establishments] [online]. [Access 20.07.2020]. Available at: http://extwprlegs1.fao.org/docs/pdf/alg41379.pdf

GRAYSON R.B., MOORE I.D., MCMAHON T.A. 1992. Physically based hydrologic modeling: 2. Is the concept realistic?. Water Resources Research. Vol. 28 p. 2659–2666. DOI 10.1029/92WR01259.

KASHANI M.H., GHORBANI M.A., DINPASHOH Y., SHAHMORAD S. 2014. Comparison of Volterra model and artificial neural networks for rainfall–runoff simulation. Natural Resources Research. Vol. 23 p. 341–354. DOI 10.1007/s11053-014-9235-y.

LEVENBERG K. 1944. A method for the solution of certain non-linear problems in least squares. Quarterly of Applied Mathematics. Vol. 2 p. 164–168. DOI 10.1090/QAM/10666.

LIN G.-F., CHEN G.-R. 2008. A systematic approach to the input determination for neural network rainfall–runoff models. Hydrological Processes. Vol. 22 p. 2524–2530. DOI 10.1002/hyp.6849.

MACHADO F., MINE M., KAVISKI E., FILL H. 2011. Monthly rainfall–runoff modelling using artificial neural networks. Hydrological Sciences Journal. Vol. 56 p. 349–361. DOI 10.1080/02626667.2011.559949.

MARQUARDT D.W. 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics. Vol. 11 p. 431–441. DOI 10.1137/0111030.

MINNS A.W., HALL M.J. 1996. Artificial neural networks as rainfall-runoff models. Hydrological Sciences Journal. 41 p. 399–417. DOI 10.1080/02626669609491511

NASH J.E., SUTCLIFFE J.V. 1970. River flow forecasting through conceptual models part I – A discussion of principles. Journal of Hydrology. Vol. 10 p. 282–290. DOI 10.1016/0022-1694(70)90255-6.

RAJURKAR M.P., KOTHYARI U.C., CHAUBE U.C. 2002. Artificial neural networks for daily rainfall–runoff modelling. Hydrological Sciences Journal. Vol. 47 p. 865–877. DOI 10.1080/02626660209492996.

RANDRIANARIVONY R.N., LAURET P., RANDRIAMANANTANY Z.A., GATINA J.C. 2009. Modélisation du régime annuel des petites rivières en vue d'installation de microcentrales hydroélectriques [Modeling of the annual regime of small rivers with a view to installing micro hydroelectric power stations]. Afrique Science: Revue Internationale des Sciences et Technologie. No. 05 p. 39–49.

REZAEIANZADEH M., STEIN A., TABARI H., ABGHARI H., JALALKAMALI N., HOSSEINIPOUR E.Z., SINGH V.P. 2013. Assessment of a conceptual hydrological model and artificial neural networks for daily outflows forecasting. International Journal of Environmental Science and Technology. Vol. 10 p. 1181–1192. DOI 10.1007/s13762-013-0209-0.

SRINIVASULU S., JAIN A. 2006. A comparative analysis of training methods for artificial neural network rainfall–runoff models. Applied Soft Computing. Vol. 6 p. 295–306. DOI 10.1016/j.asoc.2005.02.002.

VIDYARTHI V.K., JAIN A., CHOURASIYA S. 2020. Modeling rainfall-runoff process using artificial neural network with emphasis on parameter sensitivity. Modeling Earth Systems and Environment. Vol. 6 p. 2177–2188. DOI 10.1007/s40808-020-00833-7.

VILANOVA R.S., ZANETTI S.S., CECILIO R.A. 2019. 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, 105080. DOI 10.1016/j.compag.2019.105080.

YASEEN Z.M. 2015. Artificial intelligence based models for stream-flow forecasting: 2000–2015. Journal of Hydrology. Vol. 530 p. 829–844. DOI 10.1016/j.jhydrol.2015.10.038.