EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology 2024, 25(9), 68–80
https://doi.org/10.12912/27197050/190235
ISSN 2299–8993, License CC-BY 4.0

Received: 2024.06.09
Accepted: 2024.07.10
Published: 2024.08.01

Hydrological Forecasts Modeling Using Artificial Intelligence and Conceptual Models of KébirRhumel Watershed, Algeria

Khaldi Ramzi[1,2*], Marouf Nadir[2], Bouziane Mohamed Tewfik[1], Djafer Khodja Hakim[3]

[1] Department of Hydraulic, Faculty of Technology, University of Mohamed Khider, Biskra, 7000, Algeria
[2] Laboratory of Ecology and Environment, University of Larbi Ben M'hidi, Oum El Bouaghi, 4000, Algeria
[3] Institute of Technology, Water Engineering Department, University of Akli Mohand Oulhadj, Bouira, 10000, Algeria
* Corresponding author's e-mail: h.djaferkhodja@univ-bouira.dz

ABSTRACT

This study models the rainfall-runoff relationship in the Kebir-Rhumel River watershed in the Constantine Highlands, Algeria, using data from three concomitant rainfall and hydrometric stations. Statistical tests confirmed the absence of breaks in the series. We applied four conceptual models (GR4J, IHAC6, MORDOR, TOPMO8) and neural network models (RNN, NARX, LSTM) over three- and ten-year periods. Among the conceptual models, GR4J provided the best fit, highlighting the non-stationary nature of the relationship. The PMC neural network model performed well over three years but was less effective over ten years due to low flow influence. Notably, the NARX-RNN and RNN-LSTM models showed excellent predictive accuracy, with NARX-RNN perfectlycapturing flow dynamics and RNN-LSTM achieving minimal RMSE and high correlation coefficients. This study lies the comparative analysis of conceptual and neural network models, specifically the NARX-RNN and RNN-LSTM models, which have not been extensively applied in this context. This research fills the gap in understanding the effectiveness of neural network models in modelling non-stationary rainfall-runoff relationships in the region.

Keywords: rainfall-runoff, prevision, Kalman filter, artificial neurons networks, nonlinear autoregressive exogenous model, long short term memory.

INTRODUCTION

The complexity, multi-variability and nonlinearity that encompass the state transformation processes of the hydrological cycle at the watershed scale leads to a difficult prediction of the hydrometric part to be determined especially for ungauged watersheds.

Many developed models and their combinations is constantly increasing. Several interesting initiatives of application on these models were carried out on the Algerian catchment areas, where the rainfall is the runoff generator. Grouped into two climates, the high precipitation areas, such as the northern coastal regions and mountainous areas, characterized by a strong tendency to runoff, in contrast, the southern and southeastern Algerian arid and semi-arid regions have low rainfall and where runoff tends to be minimal [Amireche et al., 2018; Marouf and Remini, 2019; Zeyneb et al., 2022].

In general, the rainfall-flow relationship in Algerian watersheds is complex and varies according to location and environmental conditions. Among these attempts, many researchers have described a study inspired by the Soil Conservation Service (SCS) production function to develop a model to predict annual flows in Algerian northern region. Based on balance analysis that regulates rainfall-runoff-transport process, scientific works on sediment transport in several watersheds have been extended using statistical and mathematical models[Fartas et al., 2017; Tamrabet et al., 2019; Batout et al., 2022].They have presented the use of different conceptual, Neuro-Fuzzy models with Kalman filter integration to model this relationship at different time steps. Thus, research study of Aoulmi et al., has give satisfactory results using consists in modeling the transformation of rainfall into flows on Seybouse basin, north of Algeria, through different statistical methods and hydrological models, as RNA with integration of modern heuristic optimization approaches and deep learning of a Convolutional Neural Network (CNN) [Aoulmi et al., 2021; 2023].

This paper aims to provide the most relevant model for Kebir-Rhumel watershed covering the spatio-temporal variation of the rainfall-flow relationship, based on the used model improvement either conceptual or neural network. Indeed, the GR4J, IHAC6, MORDOR, and TOPMO8 models could not consider the chronological succession of phenomena.

According to Nash and Sutcliffle, operational flood forecasting requires the hydrological model, in addition a method for continuous correction of the forecast from the observed error of the first forecast [Duc and Youhie, 2023].

Also, there are many rainfall-flow models on a daily scale, either conceptual or mathematical (physical, black-box), giving the impression that this relationship is well determined. While theirapplication has shown that they are difficult to calibrate and the results are unsatisfactory. Among the used models, the conceptual and the physical models are the two best to understand the rainfall runoff process, although they require more watershed parameters, which also contain very complex relationships to construct these models. Furthermore, it is difficult to obtain these parameters, and black models have been increasingly highlighted in recent years.

To overcome this issue, the GR4J model, ranked better than its predecessors, is combined with the implementation of non-linear Kalman Filter. Thisimplementation generates the GR4J model with Particle Filter (GR4JPF) and the GR4J model with Ensemble Kalman Filter (GR4JENKF), where, its principle consists perturbationscreation of based on statistics in precipitation and evapo-transpiration inputs to generate different members of the ensemble as a term of uncertainty propagation.

We started using black box models in this work via the PMC model, applied over the two periods over the study area. However, a PMC may not predict the dynamic hydrological process. To improve the dynamics of simulated flows through the use of RNN based on the NARX model, which can be effectively applied to long-term forecasting of a non-linear time series for flow forecasting and prediction [Zheng et al., 2022], and an LSTM network that has a tendency to do better on a volatile time series with more than one stationary component [Li et al., 2022; Xu et al., 2022]. The two models were built using the same data as those used in the first part of daily precipitation, daily evapotranspiration determined, the evaluation is therefore established on the basis of Statistical indices of Nash - Sutcliffe efficiency (NSE), correlation coefficient (R) and root mean square error (RMSE).

PRESENTATION OF THE STUDY AREA

Centered between 7° longitude East and 36° latitude North, the Kebir-Rhumel watershed is one of the great basins of Algeria, covering an area of 8815 km², located largely on the southern slope of the Tellian bulge (Fig. 1**)**. It represents as such an intermediate area between the Tellian domain with a very strong mediterranean influence in the north and the domain of the high plains with a strong continental influence in the south.

Composed of seven hydrographic entities, the study area focuses on three watershedsareas which are Ain-Smara with altitudes up to 1462 m and an area equal to 2169 km^2, Grarem as intermediate of an area of about 5400 km^2 and a relief that reaches 1444 m, and El-Ancer of a maximum altitude of 1729 m of the coastal area hasa seafront of about 7 km^2.

DATA AND METHODS

All the basic rainfall and hydrometric data is established from the National Water Resources Agency in Algiers. The data provide the recorded values of rainfall and average daily flow of six simultaneous stations (Table 1). In the statistical analysis of meteorological series, the most commonly used methods are tests concerning the randomness of the series.

They are more particularly adapted to breaks detection in a time series.

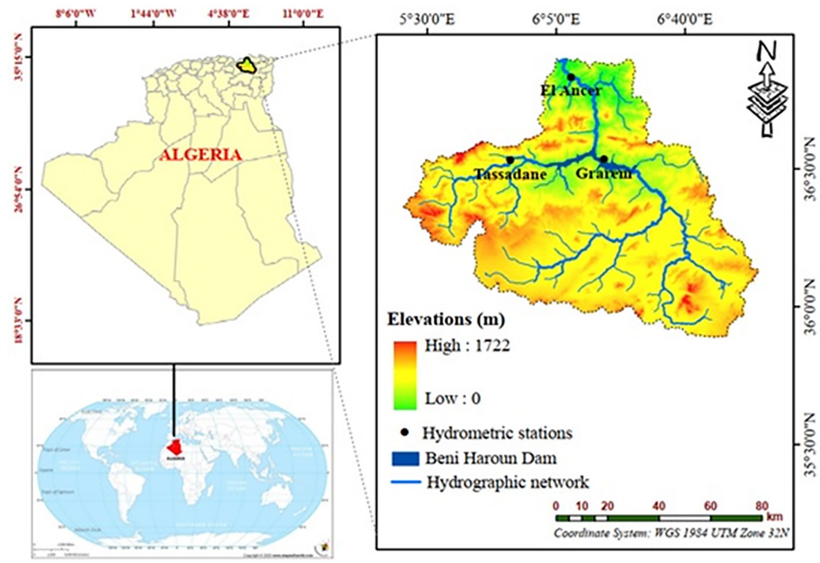

Figure 1. Kebir-Rhumel watershed and study area sub-basins (SB)

Table 1. Statistic characteristics of stations data

Station	Ain Smara			Grarem			El-Ancer		
Code	100701			100601			100109		
Parameter	P	E	Q	P	E	Q	P	E	Q
Mean	2.32	2.803	1.04	2.670	2.803	5.029	3.263	2.803	27.66
Median	0.11	2.457	0.30	0.106	2.457	2.240	0.161	2.457	8.40
Std-deviation	5.42	1.658	6.55	6.132	1.658	12.165	7.313	1.658	80.70
Variance	29.42	2.749	42.86	37.602	2.749	147.984	53.479	2.749	6512.87
Min value	0.00	0.386	0.00	0.000	0.386	0.120	0.000	0.386	0.19
Max value	70.52	7.078	200.00	72.784	7.078	213.373	131.124	7.078	2529.00

HYDROLOGICAL MODELING OF THE RAINFALL-FLOW RELATIONSHIP

Conceptual models presentation

In our study, four hydrological conceptual models were chosen, in order to obtain satisfactory results for the study area. The chosen models are: GR4J, IHACRES, MORDOR, and TOPMO8 MODEL.

GR4J model

Inour study, the 4-parameter daily rural engineering model was used as hydrological model. GR4J is a simple hydrological conceptual model with two tanks: the first is the production tank, and the routing tank.

IHACRES model

The IHACRES model has three conceptual storage models; The Nonlinear Module that determines effective rainfall and two parallel storage models that transfer effective rainfall to stream-water.

MORDOR model

The main hydrological processes of the model: evapo-transpiration, runoff, infiltration, groundwater discharge, accumulation and snowmelt. Exchanges between the different stocks and with the atmosphere and the river are regulated through simple but parameterizable formulations [Coron, 2013].

TOPMO8 model

TOPMO8 Model describes the hydrological response related to the watershed topographic characteristics, including slope and contributing area, as well as the rainfall distribution in space and time. It considers the non-linear relationships between rainfall, soil moisture and flow generationand assumes that the watershed water-balance is controlled by the water Table position related to the land surface. The four used models are summarized in the Table 2.

GR4J model and Kalman Filter "KF"

The use of updating methods is tolerated to allow the applied model to be as much as possible in conformity with the reality of the observed flows at the forecast time. Four main types of updates exist. Theyconsistrespectively in:

- a correction on the inputs (rainfall, evapo-transpiration) of the model (update of inputs).
- a correction on the internal variables (level in the tanks) of the model (update of the model states);
- a correction on the model parameters (update of the parameters);
- a correction of the model errors understood as the difference between observed and simulated flows.

In this perspective, Kalman filter is introduced in implementation with the GR4J conceptual model. Despite its interesting characteristics, Kalman filter remains a linear method with the inherent drawbacks of this method type. A variant of Kalman filter, known as Extended Kalman Filter (EKF), allows to model non-linear relationships, as the rain-flow relationship. The joint use of GR4J and EKF aims to combine the advantages of both. It takes advantage of the continuous operation of the GR4J and the real-time adaptability of the KF.

The obtained models are a reformulation of the GR4J conceptual model into a system of dynamic state equations by implementing the (EnKF) and (PF) procedure [Coron et al., 2017]. With a direct perturbation of hydrological variables, the state variables therein are the water level in the production and routing tanks and finally the flow measurements therein are the output variables.

In the EnKF prediction step, the nominal value of rainfall and evapo-transpiration is disturbed by a multiplier as follows:

$$P_t^i = P_t \times e_t^i \tag{1}$$

with

$$e_t^i = \exp(Z_t^i) \sim \log N(1, \sigma_e) \tag{2}$$

where: P_t is the rainfall measured at t time, and i is the multiplier for the i[th] member at t time, P^t_i is the disturbed rainfall for the i[th] member at t time and $z_t^i \sim N(m_z, \sigma_z^2)$.

Calculation of the covariance matrices is as follows:

$$B_t = \frac{1}{(q-1)}\left(X_t^b - \bar{X}_T^b\right)\left(X_t^b - \bar{X}_T^b\right)^T \tag{3}$$

$$R_t = \frac{1}{(q-1)}(Y_t - \bar{Y}_t)(Y_t - \bar{Y}_t)^T \tag{4}$$

where: B_t – draft covariance matrix (dimension $N \times N$); X_t^b – draft, a priori, state vector of the model (dimension $N \times q$); $\bar{X}_T^b$ – average of the state vector ensemble (dimension $N \times 1$); R_t – covariance matrix of the observations (dimension $M \times M$); y_t – vector of observations (dimension $M \times q$); $\bar{y_t}$ – average of the ensemble of observations (dimension $M \times 1$); q – size of the ensemble; N – number of state variables; M – number of observations; t – time step.

The Kalman gain (K_t) is then calculated from these covariance matrices and an observation operator (H_t):

$$K_t = (H_t B_t + R_t)^{-1} \tag{5}$$

EnKF requires calculation of a Kalman gain, which is then used as a weight to set a compromise between a model simulation from a particular state vector (a member of the ensemble) and the observed flow value, to be added to the state vector to form the analysis (X_t^a) of dimension $N \times q$:

$$X_t^a = X_t^b + K_t(y_t - H_t X_t^b) \tag{6}$$

The not updated model state variables are considered as main difference between the PF method and other methods. Among them, we based on the probability distributions that change over time**.** The particle filter advantages present the probability densities in a complete way (not limited to the first two static moments) unlike the EnKF and to consider strong non-linearity.

Table 2. Models summary and design

ID	Model	Number of optimized parameters	Number of storages
1	GR4J	4	2
2	IHAC	6	3
3	MORD	6	4
4	TOPMO8	8	3

Artificial neurons networks

Artificial neurons networks (ANN) are mathematical structures inspired by the biological neural networks. A neural network is a type of machine learning model. It is composed of layers of interconnected nodes, or neurons, that process information by applying a set of mathematical operations to inputs. Neurons in each layer communicate with neurons in adjacent layers via a set of weights, which are adjusted during training to optimize the network's ability to learn from the inputs [Vanhoucke et al., 2011].

Static neurons networks (Multi layers Perceptron)

Multi-layers perceptron (MLP) is a neural network consisting of one input layer, one or more layers of nonlinear processing. An important first step in building MLP models is to determine the best architecture, as well as the epoch's number to use for MLP training. The MLP hidden neurons number was set to vary between one and a maximum number defined as "one plus three times the number of input variables". This is a variant of the formula proposed.

The MLP consists of four layers including 30 neurons for the input layer, 25 and 20 consecutive neurons for the two hidden units, called hidden layers, and one final layer of processing units, called the output layer. Each unit in a hidden layer is typically connected to all units in the previous and subsequent layers. The MLP layers can be thought of as a sequence of transformations that map input data to a desired output [Géron, 2022].

Determination of network architectureand formation times

Layers for Ain-Smarawatershedand three layers of 25, and 20 neurons for the two watersheds of Grarem and El-Ancer. Rainfall data is also critical, as it provides data to the network about rapid increases in flow, while evapo-transpiration indicates flow attenuation, and since the goal is to predict flow at the daily time step for model output, the output layer will consist of only a single neuron, with Levenberg-Marquardt as the training algorithm and an epoch number of 1000.

Dynamic neuron networks

One of the main advantages of dynamic neuron networks (RNN) in hydrology is their ability to learn from data and adapt to changing environmental conditions. This leads them to be particularly useful to predict hydrological processes under changing climatic conditions, where traditional models may not be able to capture the complex interactions between different variables [Shao et al., 2022]. RNNs are designed to recognize the sequential characteristics of a datum and the using patterns to predict the next likely scenario.

RNN NARX

The NARX network is a dynamic recurrent network with feedback connections enclosing several layers of the network. The NARX model is based on the linear ARX model, which is commonly used in time series modeling. The defining Equation of NARX model is:

$$y(n+1) = f[y(n),...,y(n-d_y+1); u(n),u(n-1),...,u(n-d_u+1)] \tag{7}$$

where: $u(n)$ m – model input at discrete time step n; $y(n)$ – model output at discrete time step n; $d_u \geq 1$ – Input memory order; $d_y \geq 1$ – output memory order.

In this work, precipitation and evaporation are selected as input parameters because runoff is directly related to net precipitation and discharge. Note also that full consideration of the surface runoff routing time and model input components, precipitation P(t), P(t–1),..., P(t), P(t−1),…, P(t−30), evapo-transpiration E(t), E(t–1),..., E(t–30) and discharge Q(t–1), Q(t–2), Q(t–3) are selected as input parameters while the output is represented by the predicted discharge Q(t). The observed rainfall, evaporation and flow data were divided into three parts: 70%, 15%, and 15% for training, checking and testing, respectively.

In this part, and from what was studied and formed above, the NARX model optimal architecture was obtained via 2 input variables, 1hidden layer, 30 neurons, 1 thirty-day time period and 1 output target. The NARX model is optimized based on Levenberg-Marquardt algorithm.

LSTM model

The LSTM evolved from RNN [Sherstinsky, 2020], which solves the problems of gradient disappearance and gradient explosion that easily occur in RNN and can store both short-term and long-term memory in the network[Karim et al., 2017]. Compared to RNN, the LSTM neural network adds a logic gate control mechanism and a state transfer unit, so that it not only preserves the correlation with time, but also increases the dependency between the remote data.

Shows the cell unit of the LSTM neural network. LSTM can effectively learn the formulas and rules data in the historical sequence data. In the research areas, the LSTM model has achieved high accuracy and it is a very effective neural network model [Sun et al., 2018].

In particular, the LSTM has two interconnected hyper-parameters that must be set together: the input sequence length (the lookback) and the hidden unit size. For this, for each LSTM in this paper, we tested combinations of the hyper-parameter variations whose values taken are – 4 (batch size (60, 72, 128 and 256)) × 4 (hidden unit size variations (50, 64, 128 and 256)) × 3 (dropout rate variations (0.25, 0.4 and 0.6)) = 48 set cases. In all these cases, the used optimizer is Adam and the learning rate is constant and equal to 10^{-3}. The 0.25 and 0.4 dropout rates did not provide any performance improvements relating the 0-dropout rate, but the 0.6 dropout rate contributed to a remarkable improvement. All results presented here correspond to the 0.6 dropout rate.

RESULTS AND DISCUSSIONS

Cross-referencing the used conceptual models and corresponding Nash criteria values, as well as the correlation coefficient values as expressed in Table 3, using as data the daily precipitations, evapo-transpiration. Comparison of these obtained results at the three studied watershed has led to the fact that the GR4J model presents the best performances to reproduce basic flows, but with weak quality regarding floods.

This model is considered as the most robust for semi-arid regions and fluctuating hydrological regime for Ain-Smara and Graremwatersheds, over a limited period of 3 years, while for the coastal El-Ancerwatershed, the simulation results confirm that the semi-distributed MORDOR model outperforms the other models. Show that small basins benefit less from the semi-distribution, this could justify the GR4J conceptual model superiority.

For Ain-Smarawatershed, the resulting modeled flow time series appears to underestimate the frequency of fast flow events (Fig. 2). Despite the relatively good values of the performance indicators for Graremand El-Ancer watershed, it is observedthat the peak flows are not correctly represented. In particular, in some cases, the modeled spikes occur when there are no observed peak flows (Fig. 3 and Fig. 4).

Thus, for the simulation results with both Kalman filters (EnKF, PF) versions, in both Ain-Smara, Graremwatershed and El-Ancer watershed, it is noticed that the flows dynamics is improved (Fig. 5, Fig. 6 and Fig. 7), andnophase shift neithercreation of new peaks have appeared. On the other hand, the flood flow values were not reproduced, which can be

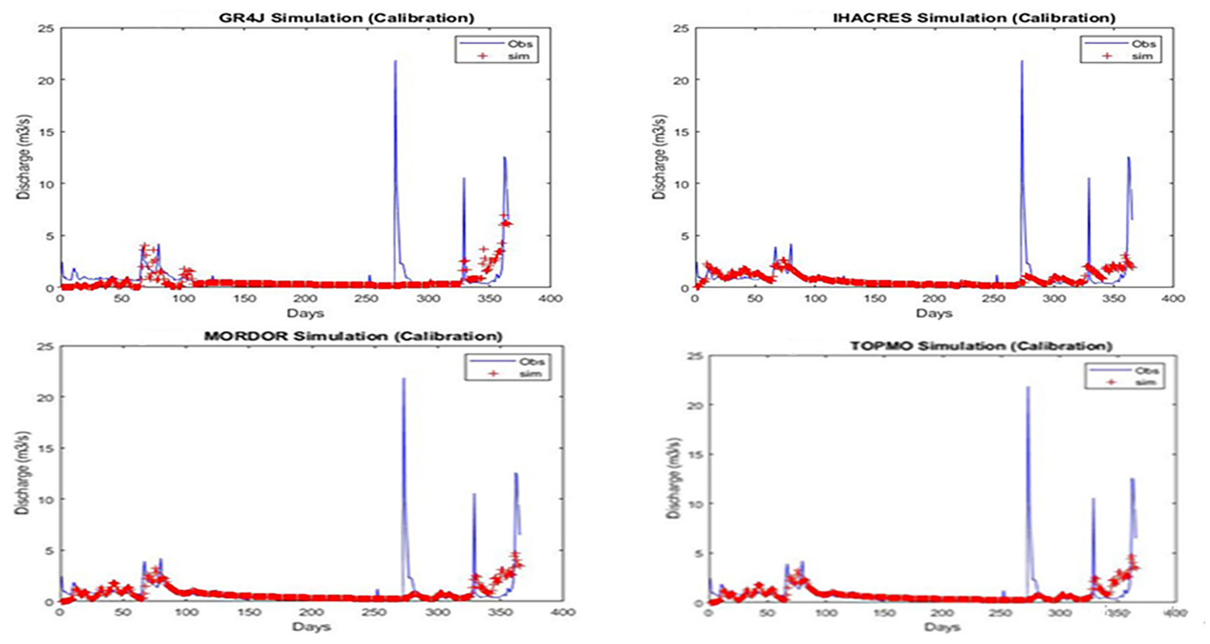

Figure 2. Correlation and variability of observed and simulated flows via

Table 3. Obtained values of the efficiency criteria

ID	Model	Basins											
1	Conceptual model	AIN SMARA Station				GRAREM Station				EL- ANCER Station			
		Critères of performances											
	Period 03 years	NSE (%)	R			NSE (%)	R			NSE (%)	R		
			Calibration		Test		Calibration		Test		Calibration		Test
1.1	GR4J	19.08	0.457		0.710	82.00	0.905		0.682	31.87	0.574		0.624
1.2	IHACRES	12.15	0.340		0.764	34.57	0.587		0.630	60.84	0.763		0.710
1.3	MORDOR	15.71	0.391		0.783	53.37	0.730		0.680	69.71	0.827		0.740
1.4	TOPMO8	-408	0.048		0.401	-23.42	-.0209		0.475	58.56	0.797		0.655
2	Conceptual hybrid models	NSE (%)				NSE (%)				NSE (%)			
2.1	GR4J PF	68.38				51.57				42.78			
2.2	GR4J ENKF	52.50				-12.53				-321.36			
3	Neural models												
	Period 03 years	NSE (%)	R			NSE (%)	R			NSE (%)	R		
			Training	Validation	All		Training	Validation	All		Training	Validation	All
3.1.1	MLP	83.76	0.9323	0.7105	0.9157	87.45	0.9468	0.5786	0.9355	97.33	0.9905	0.9964	0.9910
3.1.2	NARX	93.62	1.00	0.768	0.9962	96.65	1.00	0.7131	0.9839	95.80	1.00	0.8509	0.9886
3.1.3	SLTM	RMSE			0.88	RMSE			0.962	RMSE			0.93
		0.0943				0.50				3.29			
	Period 10 years	NSE (%)	R			NSE (%)	R			NSE (%)	R		
			Training	Validation	All		Training	Validation	All		Training	Validation	All
3.2.1	MLP	63.37	0.8475	0.8516	0.8259	25.5	0.5093	0.4070	0.5036	48.20	0.6608	0.8327	0.6977
3.2.2	NARX	97.70	0.997	0.9112	0.9627	86.75	0.9978	0.7035	0.9419	84.00	0.9992	0.6472	0.9300
3.2.3	SLTM	RMSE			0.85	RMSE			0.90	RMSE			0.90

justified by the underestimation of the unit hydrograph X_4. However, [Pauwels et al., 2013] mentioned that the EnKF was not well suited when a unit hydrograph was used due to the error of the time propagating hydrograph.

Thus, referring to Nash index values, the best results are those obtained by GR4J and PF (Crit.NSE = $St_{Ain\text{-}Smara}$ = 0.6838, St_{Grarem} = 0.5157 and $St_{El\text{-}Ancer}$ = 0.4278. compared to hybrid model GR4JEnKF (Crit.NSE = $St_{Ain\text{-}Smara}$ = 0.5250, St_{Grarem} = -0.1253 and $St_{El\text{-}Ancer}$ = -3.2136).

Even for the PF model, there is no constraint related to states distribution, contrary to EnKF-model requires a Gaussian distribution.

Regarding the 2nd concept when using RNA at Ain-Smarawatershedover three years period, the obtained results are satisfactory with a correlation coefficient and Nash number of about 0.91 and 0.837, respectively. The simulated flows are criticized for reproducing all values except three spikes (Fig. 8).While over the 10-year period, a clear degradation is observed on the flow dynamics but with a correlation coefficient and a Nash index of 0.825 and 0.63, respectively. The same observation is obtained for the other watersheds with efficiency indices of Grarem (R = 0.93, NSE = 0.874) and El-Ancer (R = 0.99, NSE = 0.97) watershedsover three years period, and a poor-quality simulation concerning the second duration (Fig. 9 and Fig. 10). Alternatively, machine learning (MLP) based on black-box models that have a higher computational capacity, is more suitable than conceptual models for the Kebir-Rhumel watershed. NARX models performance is summarized and evaluated using the same efficiency indices NSE, CCand the graphical chronology of the simulated and recorded flows. The results showed that the accuracy of the rainfall-runoff simulation based on NARX model is better than that of MLP model in the training, validation and test periods for all three stations.A significant improvement is obtained for all the studied watersheds, over the three-year period.

- For Ain-Smara: R = 0.936 and NSE = 0.996.
- ForGrarem: R = 0.966 and NSE = 0.98,
- ForEl-Ancer: R = 0.95 and NSE = 0. 98.

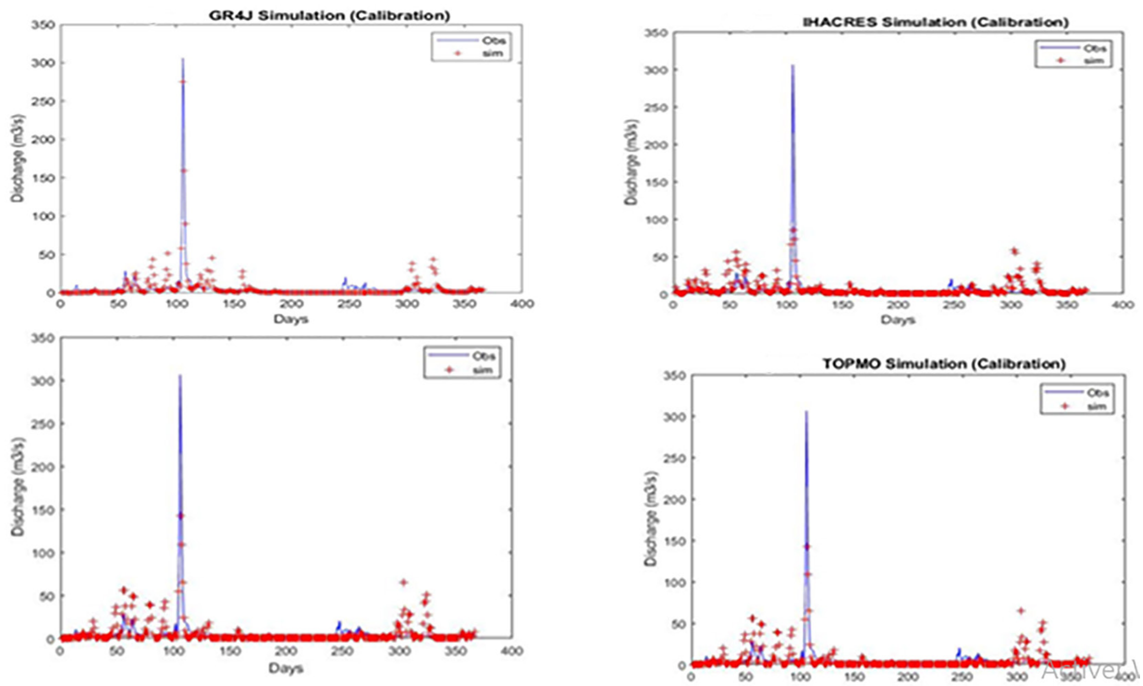

Figure 3. Correlation and variability of observed and simulated flows via (GR4J, IHAC6, MORDOR, and TOPMO8) models on Grarem SB

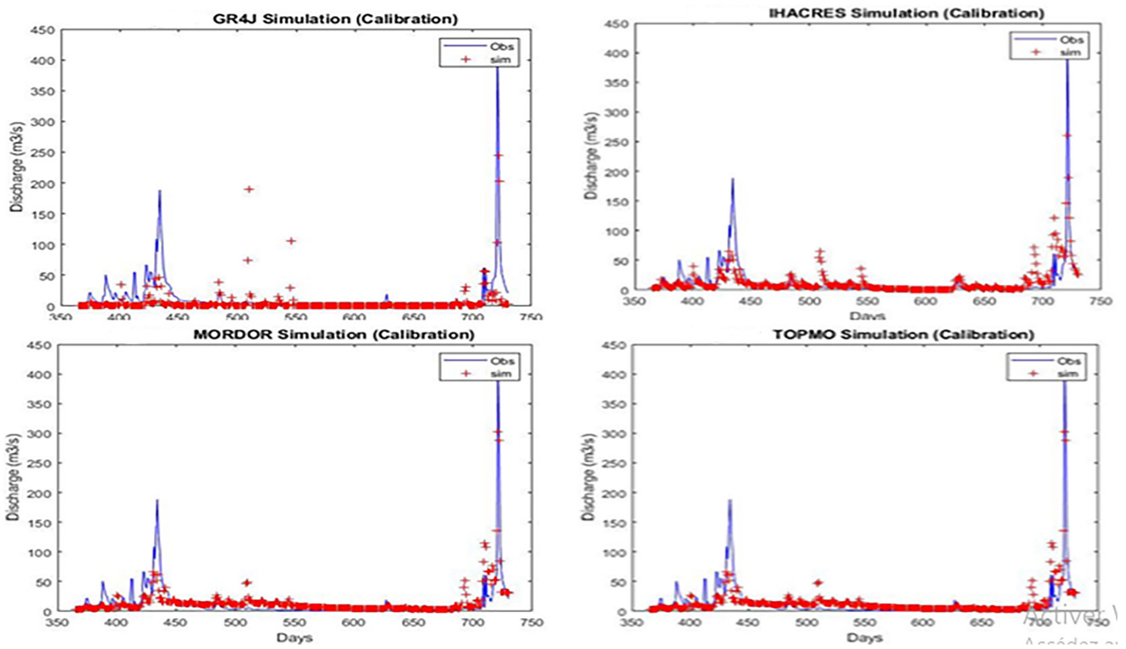

Figure 4. Correlation and variability of observed and simulated flows via (GR4J, IHAC6, MORDOR, and TOPMO8) models on El-Ancer SB

Also, the NARX modelpower was proven over the ten-year period with efficiency indices

- Ain-Smara: R = 0.977, and NSE = 0.96.
- Grarem: R = 0.867, and NSE = 0.94
- El-Ancer: R = 0.84, and NSE = 0.93.

Then the NARX model is able to simulate the dynamic rainfall-flow process very well, although there is a small underestimation of the peak flow (Fig. 11, Fig. 12, and Fig. 13). The underestimation of the peak flow may be caused by the sigmoid activation function.

For sigmoid functions, the absolute value of the independent variable increases and the slope gradually decreases to zero, indicating that there

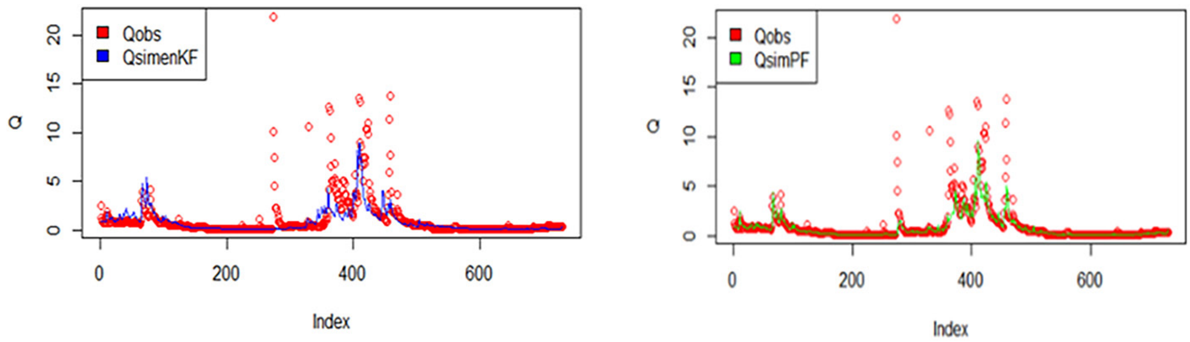

Figure 5. Variability of observed and simulated flows via GR4J and Kalman Filter hybrid models on Ain-Smara SB

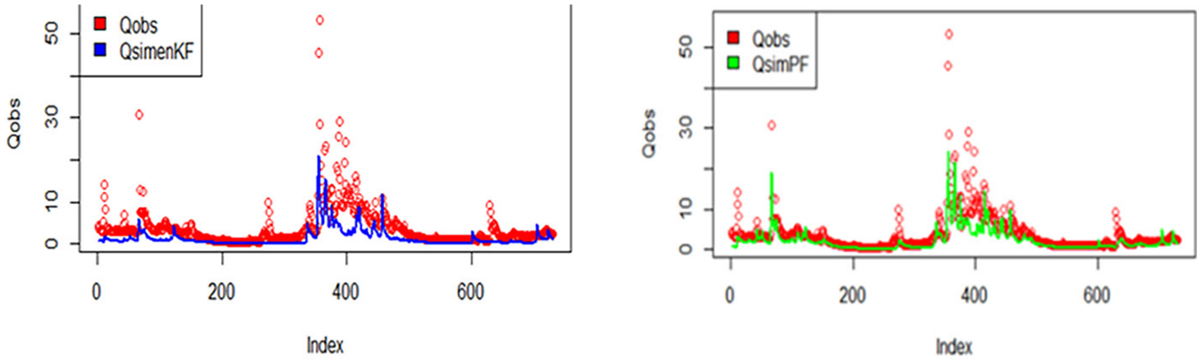

Figure 6. Variability of observed and simulated flows via GR4J and Kalman Filter hybrid models on Grarem SB

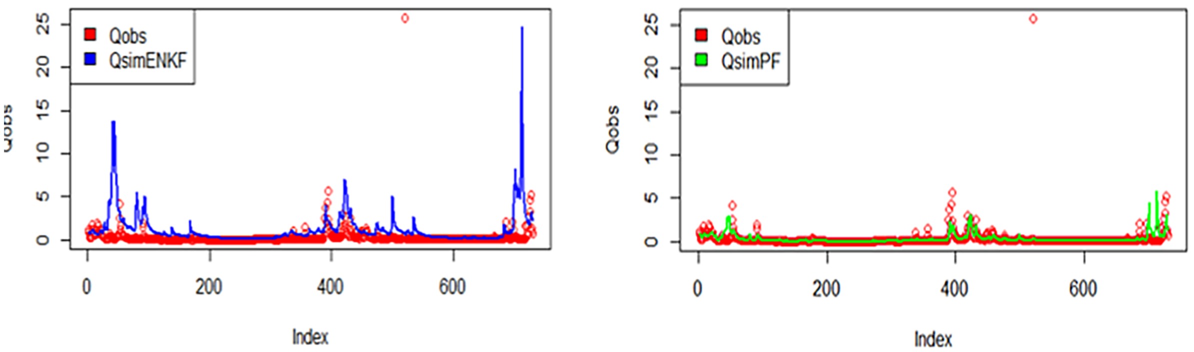

Figure 7. Variability of observed and simulated flows via GR4J and Kalman Filter hybrid models on El-Ancer SB

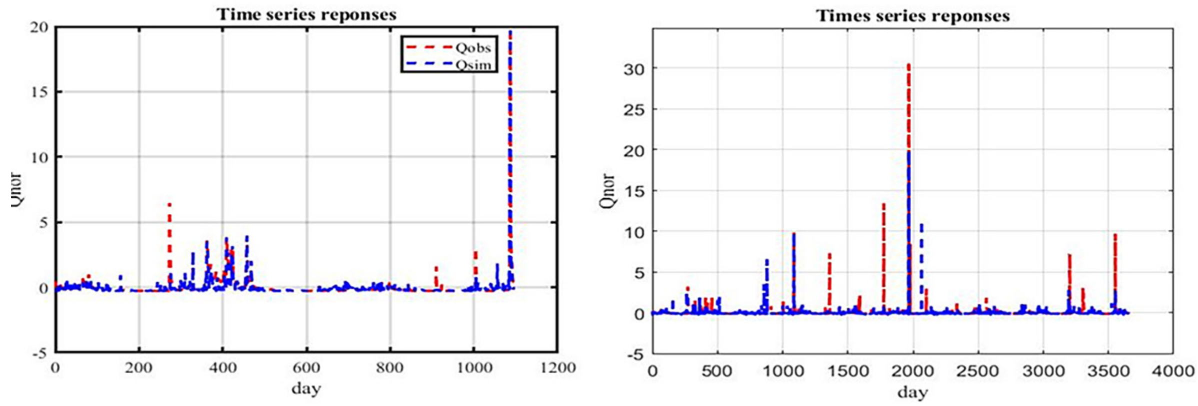

Figure 8. Presentation of MLP model efficiency indices of Ain-Smara SB

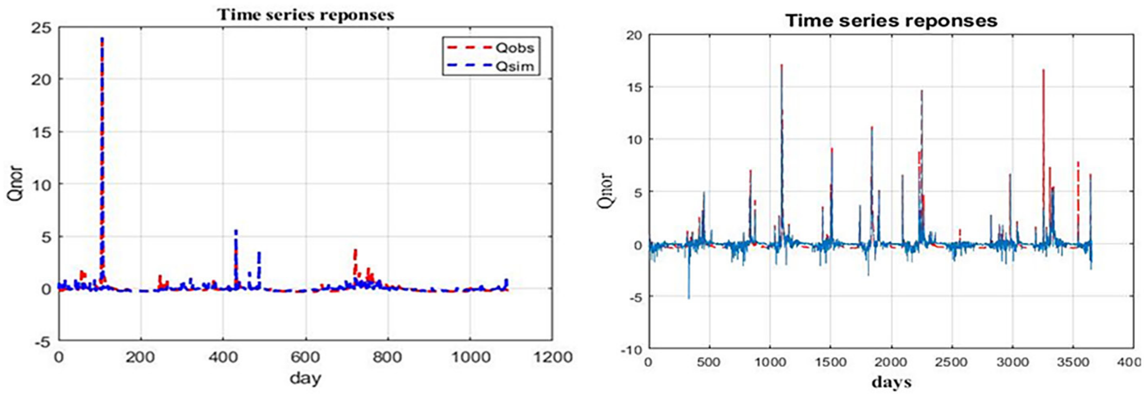

Figure 9. Presentation of MLP model efficiency indices of Grarem SB

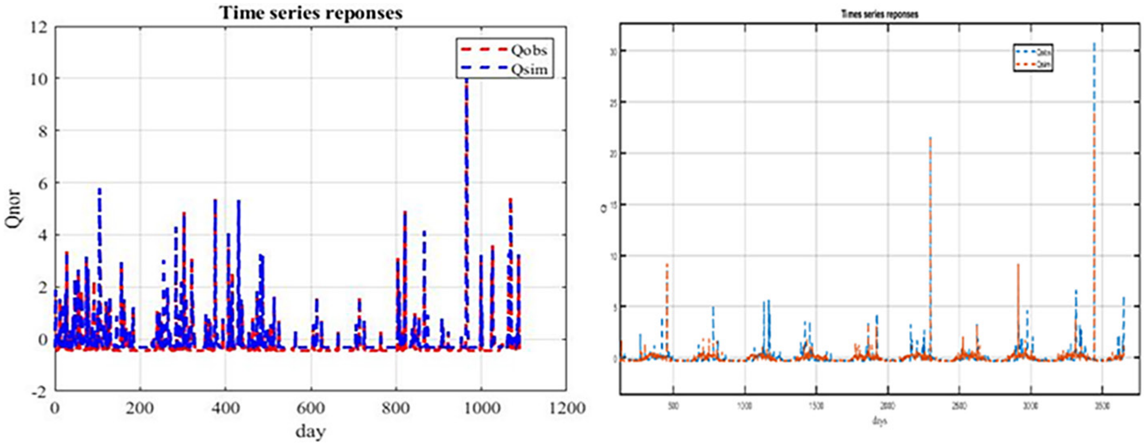

Figure 10. Presentation of MLP model efficiency indices of El-Ancer SB

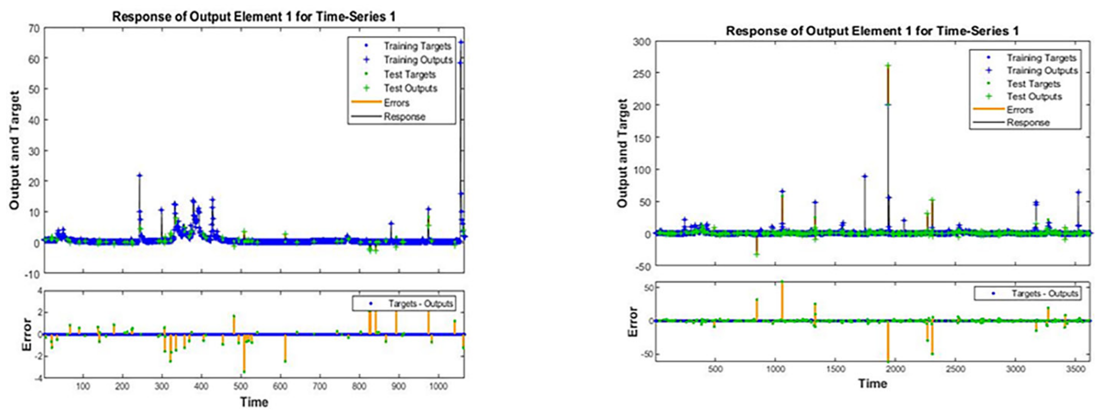

Figure 11. Presentation of NARX model efficiency indices of Ain-Smara SB

is an obvious discrepancy when a higher or lower flow is transformed through activation functions.

Compared to MLP network structures, NARX network shows a faster convergence. Regarding the last phase of this paper, LSTM model application also resulted in the best predictions which are defined by a number of hidden layer nodes, batch size and epoch number of 50, 60 and 60, respectively. The prediction results are shown in Figure 14, 15, 16. From a visual standpoint, the prediction results via LSTM neural network have improved the accuracy. Again, the modeled time series appears to perfectly capture the timing of the observed peak flows.

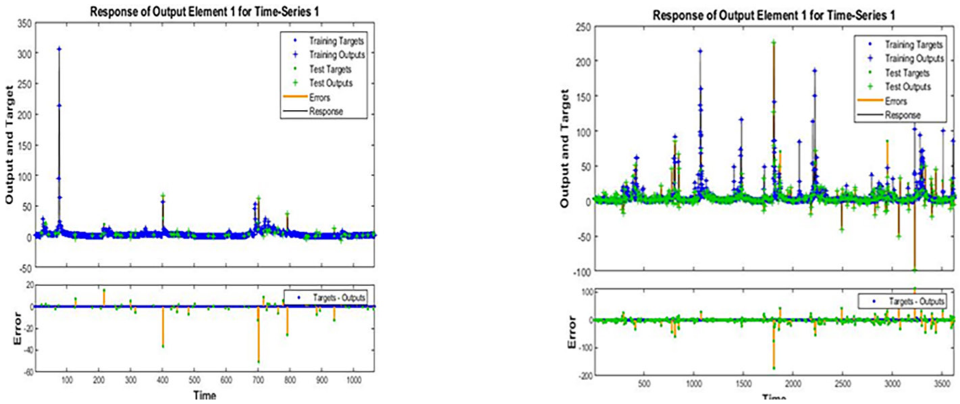

Figure 12. Presentation of NARX model efficiency indices of Grarem SB

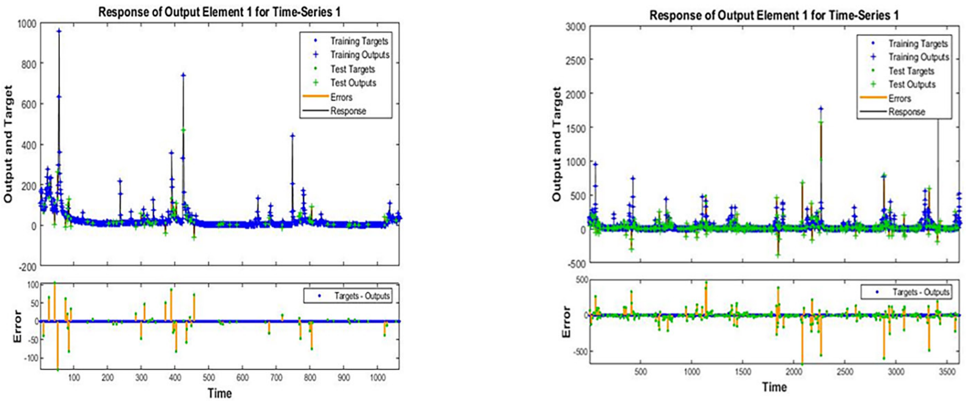

Figure 13. Presentation of NARX model efficiency indices of El-Ancer SB

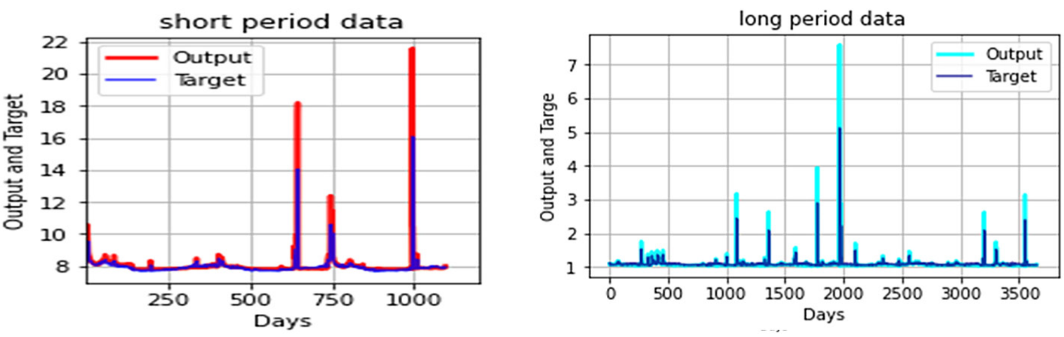

Figure 14. Presentation of SLTM model efficiency indices of Ain-Smara SB

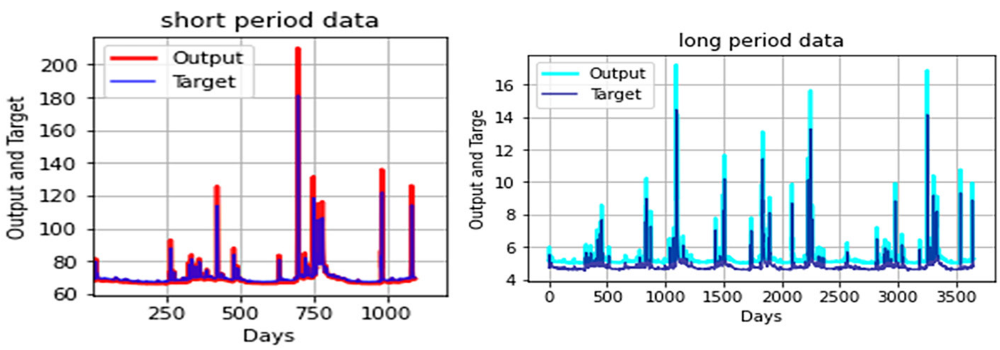

Figure 15. Presentation of SLTM model efficiency indices of Grarem SB

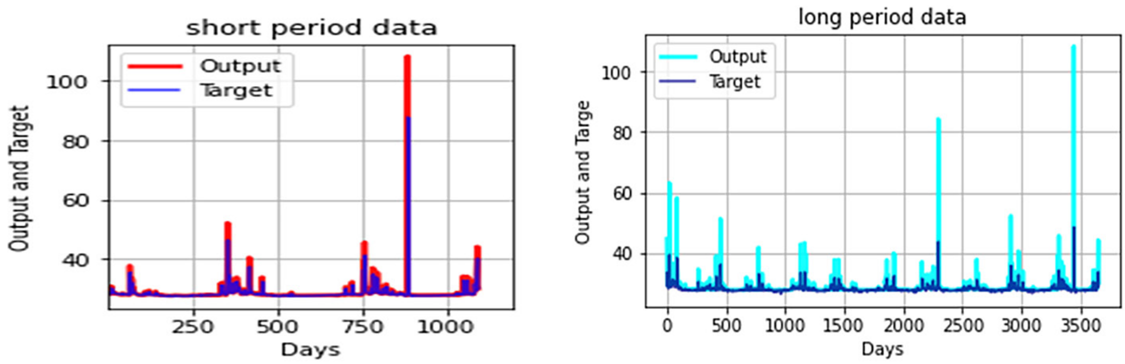

Figure 16. Presentation of SLTM model efficiency indices of El-Ancer SB

However, the results show an underestimation in the reconstruction of the maximum daily flows but the prediction overall accuracy is very satisfactory, as confirmed by the obtained performance indicators values. Thus, the model efficiency can still be considered as perfect also with respect to the simulation of the slow flow component.Moreover, RMSE values especially for Ain-Smara and Graremwatersheds related to fraction order (0.0943 and 0.5 for three years) and (0.2 and 0.48 for ten years), reveal that the error in the base flow volume evaluation decreases in absolute terms compared to calibration. For this purpose, it should be emphasized that the simulation is always carried out at daily time scale.

CONCLUSIONS

This study provides significant insights into the modeling of the rainfall-runoff relationship in the Kebir-Rhumel River watershed using both conceptual and machine learning models. Several novel contributions and findings have emerged from this research:

- Neural model efficacy: The study demonstrates that neural network models, specifically NARX-RNN and RNN-LSTM, offer superior predictive accuracy compared to traditional conceptual models and standalone neural networks. These models excel in capturing complex, non-stationary hydrological processes, which is crucial for accurate rainfall-runoff modeling in semi-arid regions.
- Conceptual model performance: Among the conceptual models tested, GR4J provided the best fit for the Kebir-Rhumel watershed over both three- and ten-year periods. The combination of GR4J with Kalman Filter techniques (ENKF and PF) further improved the model's performance, indicating that these enhancements are beneficial for more accurate hydrological predictions.
- Model Comparison: The study compares the performance of four conceptual models (GR4J, IHAC6, MORDOR, TOPMO8) and multiple neural network architectures over different periods. This comprehensive comparison establishes the relative strengths and weaknesses of each approach, providing a benchmark for future studies in similar hydrological contexts.
- Long-term Performance: The research highlights the challenges associated with long-term prediction using neural network models, particularly noting that static models like MLP struggle with watershed dynamics over extended periods. However, dynamic models such as NARX and LSTM successfully capture peak flood discharge and peak occurrence time, reflecting complex nonlinear hydrologic dynamics with high fidelity.
- Regional Application: This research fills a critical knowledge gap by applying advanced neural network models to the Kebir-Rhumel River watershed, a region where such models have not been extensively utilized. The findings provide valuable insights into the applicability and effectiveness of these models in semi-arid climates, where the runoff generation mechanism is more complicated than in humid regions.
- Evaluation Criteria: The models were evaluated using NSE, R, and RMSE criteria, which confirmed their ability to reproduce recorded observations and combined chronographs accurately. The NARX and LSTM models, in particular, demonstrated excellent performance in simulating the rainfall-runoff process, despite a slight underestimation of peak flow.

Overall, this study contributes to the advancement of hydrological modeling by validating the use of neural network models in non-stationary environments. The integration of conceptual models with Kalman Filter techniques and the application of advanced neural networks provide a robust framework for future research and practical applications in water resource management. The findings offer a comprehensive understanding of the rainfall-runoff relationship in the Kebir-Rhumel watershed, paving the way for more accurate and reliable hydrological predictions in similar regions.

Acknowledgments

We thank the national hydraulic resource agency of the eastern region, particularly of the Constantine region.

REFERENCES

1. Amireche M., Abdelmalek B., Djamel B. 2018. Modélisation de la relation pluie-débit a différents pas de temps par les modèles conceptuels, neuro-flous et par le filtre de kalman.PhD Thesis.
2. Aoulmi Y., Marouf N., Mohamed A. 2020. The assessment of artificial neural network rainfall-runoff models under different input meteorological parameters. Case study: Seybouse basin, Northeast Algeria. Journal of Water and Land Development, 50 (VI–IX), 38–47.
3. Aoulmi Y., Marouf N., Rasouli K., Panahi M. 2023. Runoff predictions in a semiarid watershed by convolutional neural networks improved with metaheuristic algorithms and forced with reanalysis and climate data. Journal of Hydrologic Engineering, 28(7), 04023018.
4. Batout S., Houichi L., Marouf N. 2022. Influence of the envelope curve on the estimate of probable maximum precipitation (PMP) in the coastal region of Algeria. Modeling Earth Systems and Environment, 8(2), 2083–2093.
5. Coron L. 2013. Les modèles hydrologiques conceptuels sont-ils robustes face à un climat en évolution? Diagnostic sur un échantillon de bassins versants français et australiens Doctorat Hydrologie, Institut des Sciences et Industries du Vivant.
6. Coron L., Thirel G., Delaigue O., Perrin C., Andréassian, V. 2017. The suite of lumped GR hydrological models in an R package. Environmental modelling & software, 94, 166–171.
7. Duc L., Yohei S. 2023. A signal-processing-based interpretation of the Nash–Sutcliffe efficiency. Hydrology and Earth System Sciences, 27(9), 1827–1839.
8. Fartas F., Marouf, N., Remini, B. 2017. Quantification du transport solide en suspension dans le barrage de Foum El Gherza–Biskra. Algerie. Journal of Water and Environmental Sciences, 1, 198–218.
9. Geron A. 2022. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc.
10. Karim F., Majumdar S., Darabi H., Chen S. 2017. LSTM fully convolutional networks for time series classification. IEEE access, 6, 1662–1669.
11. Li P., Zhang J., Krebs P. 2022. Prediction of flow based on a CNN-LSTM combined deep learning approach. Water, 14(6), 993.
12. Marouf N., Remini B. 2019. Impact study of Beni-Haroun dam on the environmental and socio-economic elements in Kébir-Rhumel basin, Algeria. Journal of Water and Land Development, 43 (X–XII), 120–132.
13. Pauwels V.R., De Lannoy G.J., Hendricks Franssen H.-J., Vereecken H. 2013. Simultaneous estimation of model state variables and observation and forecast biases using a two-stage hybrid Kalman filter. Hydrology and earth system sciences, 17(9), 3499–3521.
14. Shao Y., Zhao J., Xu J., Fu A., Li, M. 2022. Application of rainfall-runoff simulation based on the NARX dynamic neural network model. Water, 14(13), 2082.
15. Sherstinsky A. 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.
16. Sun Q., Jankovic M.V., Bally L., Mougiakakou S.G. 2018. Predicting blood glucose with an lstm and bi-lstm based deep neural network. 14th symposium on neural networks and applications (NEUREL).
17. Tamrabet Z., Marouf N., Remini B. 2019. Quantification of suspended solid transport in Endja watercourse [Dehamecha basin-Algeria]. GeoScience Engineering, 65(4), 71–91.
18. Vanhoucke V., Senior A., Mao, M.Z. 2011. Improving the speed of neural networks on CPUs.
19. Xu Y., Hu C., Wu Q., Jian S., Li Z., Chen Y., Zhang G., Zhang Z., Wang S. 2022. Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. Journal of Hydrology, 608, 127553.
20. Yunpeng L., Di H., Junpeng B., Yong Q. 2017. Multi-step ahead time series forecasting for different data patterns based on LSTM recurrent neural network. 14th Web Information Systems and Applications Conference (WISA).
21. Zeyneb T., Nadir M., Boualem R. 2022. Modeling of suspended sediment concentrations by artificial neural network and adaptive neuro fuzzy interference system method–study of five largest basins in Eastern Algeria. Water Practice & Technology, 17(5), 1058–1081.
22. Zheng Y., Zhang W., Xie J., Liu Q. 2022. A water consumption forecasting model by using a nonlinear autoregressive network with exogenous inputs based on rough attributes. Water, 14(3), 329.