Streamflow prediction using a new approach of hybrid artificial neural network with discrete wavelet transform. A case study: the catchment of Seybouse in northeastern Algeria.

TIKHAMARINE Yazid [1], SOUAG-GAMANE Doudja [2], MELLAK Soumia [3] and CHERGUI Sakina [4]

[1,2,3,4] *LEGHYD Laboratory, Civil Engineering Faculty, University of Sciences and Technology Houari Boumediene, BP 32 AlAlia, Babezzouar, 16111 Algiers, Algeria.*

E-Mail: ytikhamarine@usthb.dz

Abstract— predicting streamflow values accurately is vitally important for hydrology and hydrogeology in water resources management system. Daily and monthly streamflow prediction can help in water management domain, regulation distribution of dams and estimation of groundwater level, especially in drought and flood issues. Streamflow forecast contributes to improve long and short-term time series by using previous information, therefore a power performance model should be used to process the complex nonlinear relation between the predictor and predictive variables.

This present study investigates the performance of hybrid artificial neural network (ANN) with discrete wavelet transforms (DWT) and compared with the single model of artificial neural network (ANN) based on feed forward Back-propagation technique and Bayesian regularization algorithm. The monthly streamflow data from the Bouchegouf gauge station on Seybouse watershed (Code 14.05.01) in Algeria River is used in this study. The statistical evaluation performance criteria used are: root mean square error (RMSE), mean absolute error (MAE), Nash Sutcliffe efficiency (NSE), and correlation coefficient (R) were employed to evaluate the results performances.

The obtained results indicate that conjunction of discrete wavelet transform with artificial neural network performed better than the single (ANN) and this hybrid model could be a useful tool for solving many prediction issues.

Key-Words— discrete wavelet transform, Streamflow prediction, artificial neural network, Seybouse watershed.

I. INTRODUCTION

THIS document Streamflow forecasting is an important issue in many hydrology and water resources activities associated with management, operation and planning of water resources and dams. Currently, scientific researches on innovative systems in the fields of water resources are undertaken to quantify and understand watershed-scale hydrologic processes based on description of the previously recorded streamflow amounts.

So far, various methods have been proposed which are capable of forecasting streamflows with different accuracy levels under different conditions. The traditional regression-based approaches as stochastic and conceptual models which are unable to modeling the streamflow data accurately [1]. Thus, they exhibit a weak performance in modelling nonlinear relationship in regression problems [2].

Artificial intelligence (AI) has proved to be a very successful forecasting and prediction tool compared with traditional stochastic models [1].

The (AI) has demonstrated a high accuracy in many prediction issues concern the field of hydrology [3-4]. ANN is one of the most (AI) model used in hydrological modeling. It proved to be a very popular and successful forecasting and prediction tool.

Recently, the application of wavelet transform, for improving forecasting models accuracy, has received much attention [5-7]. All these studies showed that the conjunction of wavelet transform with artificial intelligence could perform better than the single artificial intelligence.

The purpose of this paper is to investigate the performance of conjunction wavelet-artificial neural network compared to single artificial neural network, for streamflow forecasting at the Seybouse River catchment in Algeria.

Streamflow forecasting of Seybouse River was carried out using single ANN models and hybrid wavelet-ANN models. For the latter, the effect of different vanishing moments of Daubechies wavelets on the prediction accuracy has been assessed

II. METHODS

II.1 Artificial Neural Networks (ANN)

Artificial Neural Networks can learn and estimate any complex functional relationship with high accuracy because it is one of the mathematical models that emulate the ability of the human brain to learn from experience.

ANN currently have numerous real-world applications, such as time series prediction, rule-based control and streamflow prediction [1].

The ANN has three layers; the first layer called the input layer connects with the input variables. the last layer, called the output layer, connects with the output variables and has a hidden layer in the middle contain several neurons. figure 4. More details about ANN can be found in [2]-[3].

II.2 Wavelet artificial Neural Networks (WANN)

The WANN is the conjunction of wavelet transform with artificial neural network, which decomposes an input time series into approximations and details components using wavelet transform. There are several types and decompositions wavelets; discrete (DWT) and continuous (CWT) wavelet transform [4], in addition different wavelet mothers; Haar, Daubechies, Dmeyer, Coiflets, Mexican-hat and others.

In the present study, the discrete wavelet was chosen and the Daubechies mother wavelet was selected, because it is popular wavelet and one of the most widely used in hydrological field [5]. More details about Wavelet transform can be found in [4] and [6].

II.3 Performance criteria Indicators

In this paper, the performance is examined by:

II.3.1 Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_F(i) - Q_O(i))^2} \quad (01)$$

II.3.2. Mean absolute error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Q_O(i) - Q_F(i)\right| \quad (02)$$

II.3.3. Coefficient of correlation (R):

$$R = \frac{\frac{1}{n}\sum_{i=1}^{n}(Q_O(i) - \overline{Q_O})(Q_F(i) - \overline{Q_F})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_O(i) - \overline{Q_O})^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_F(i) - \overline{Q_F})^2}} \quad (03)$$

II.3.4. Nash-Sutcliffe efficiency coefficient (NSE):

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_O(i) - Q_F(i))}{\sum_{i=1}^{n}(Q_O(i) - \overline{Q_O})} \quad (04)$$

Where n is the number of data points, $Q_F(i)$ is the forecasted value, $Q_O(i)$ is the observed value. $\overline{Q_F}$ and $\overline{Q_O}$ are the average value of the forecasted and observed respectively.

Table 1. Different input combinations

Models	INPUTS	OUTPUT
M1	Q_t	Q_{t+1}
M2	$Q_t + Q_{t-1}$	Q_{t+1}
M3	$Q_t + Q_{t-1} + Q_{t-2}$	Q_{t+1}
M4	$Q_t + Q_{t-1} + Q_{t-3} + Q_{t-4} + Q_{t-5} + Q_{t-9} + Q_{t^2-10} + Q_{t-11}$	Q_{t+1}
M5	$Q_t + Q_{t-1} + Q_{t-3} + Q_{t-10}$	Q_{t+1}
M6	$Q_t + \ldots\ldots + Q_{t-11}$ (all)	Q_{t+1}

II.1 Study area and model development

The time series of the Bouchegouf gauge station (36°, 27', 00" N, 7°, 34', 12" E) on Seybouse watershed in northeastern of Algeria is chosen in this study. The data from September 1970 to February 1988 is used for training (70%) whilst that from March 1988 to August 1995 is used for testing (30%). The observed data records were obtained from National Agency of Water Resources, Algeria (ANRH).

In the present paper, all developed models were trained using Bayesian regularization algorithm (Trainbr) to train ANN and the numbers of neurons was varied for 1 to 20 neurons for each simple (ANN) and hybrid model (WANN). Also, the time series of streamflow was decomposed using the most popular wavelet of Daubechies (db). Furthermore, a several decomposition levels of wavelet transform were tested to select the ideal level and best vanishing moments until we find the best model, which give high performance in the testing period. The different developed models according to input combinations are presented in the Table 1 using the autocorrelation function (ACF).

III. Results and Discussion

The one-step ahead forecast accuracy of the six models in preliminary testing stage (M1 to M6) was evaluated using the efficiency criteria; (RMSE, MAE, R and Nash). The model which presents the highest values of R and Nash and lower values of RMSE and MAE is the best one.

The behaviors of these parameters using ANN are presented in Table 2

Table 2. Models performance using ANN

Model	ANN in the testing period				
	Best Structure	RMSE (m3/s)	MAE (m3/s)	R (%)	Nash (%)
M1	1-10-1	2.1848	1.4108	47.62	10.95
M2	2-18-1	2.1388	1.5642	43.80	14.66
M3	3-13-1	2.1397	1.5607	44.31	14.59
M4	8-04-1	2.0397	1.2218	51.18	22.39
M5	**4-04-1**	**2.0253**	**1.2711**	**52.70**	**23.48**
M6	12-16-1	2.1139	1.4376	46.01	16.64

In additional, the best architecture of ANN was presented in the same table.

Whereas the optimum decomposition level was usually determined through Nourani formula [7]. The best vanishing moments was obtained by varying the number of vanishing moments from 1 to 20 until finding the best one based on optimum criteria measurements.

According to the obtained results, the Model M5 is the fittest model regarding to the other models with highest R and Nash (52.70% and 23.48% respectively) and lowest RMSE and MAE (2.0253 and 1.2711 m3/s respectively). The scatterplot for the best model (M5) between observed and predicted shown in figure 1.

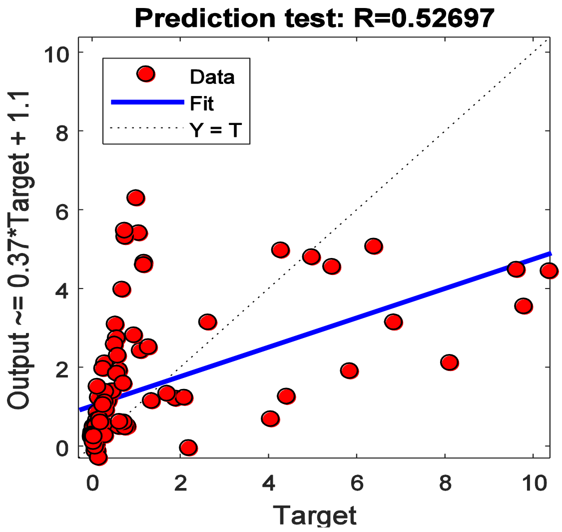


Figure 1. Scatterplot of predicted and observed streamflow values using ANN for best Model (M5).

In order to improve the prediction process using wavelet transform coupled with ANN, we use the decompositions of approximations and details as Input to the model for building a new hybrid model. The best decomposition level and the best vanishing moment for each developed model were presented in Table 3. Whilst the figure 2 shows the scatter plot, of the best model results, between observed and predicted streamflows (M3) in the testing period. It can be observed that the model M3 is the fittest model compared to the other models with highest R and Nash values (96.62% % and 93.27% respectively) and lowest RMSE and MAE (0.6008 and 0.4327 m3/s respectively).

Table 3. Models performance using hybrid WANN

Model	WANN performance in the testing period						
	Best Structure	**Best level**	**Best vanishing moment**	**RMSE (m3/s)**	**MAE (m3/s)**	**R (%)**	**Nash (%)**
M1	3-05-1	2	13	1.422	0.860	79.01	62.28
M2	6-04-1	2	15	0.646	0.445	96.07	92.22
M3	**9-08-1**	**2**	**15**	**0.600**	**0.432**	**96.62**	**93.27**
M4	24-03-1	2	13	0.739	0.470	94.83	89.80
M5	12-05-1	2	13	0.728	0.492	95.07	90.09
M6	36-06-1	2	13	0.852	0.674	94.94	90.05

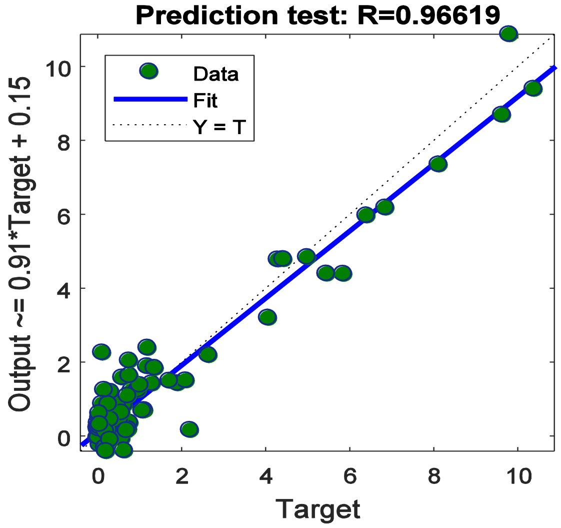
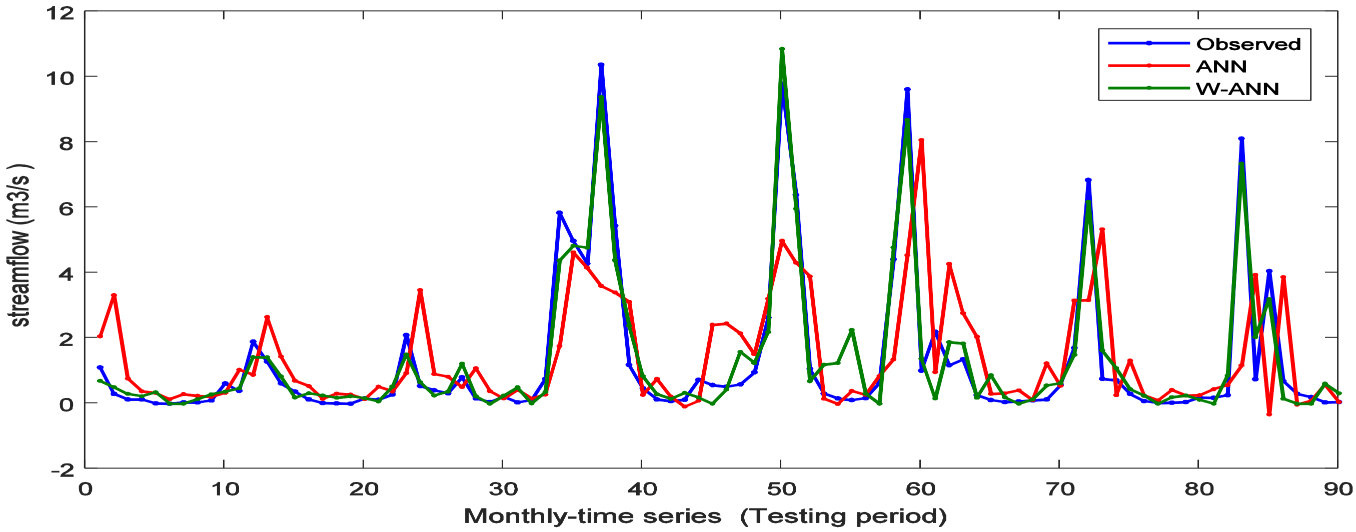


Figure 2. Scatterplot of predicted and observed streamflow using WANN for best Model (M3).

The forecasting results have all shown that WANN models are the most effective for streamflow forecasting for all inputs combinations (six different models) in terms of performance criteria, compared to single ANN. It has also an effect on the optimum number of neurons in hidden nodes (decreased in comparison to the ANN models). In addition, selection of inputs combination has a great effect on the model's accuracy. The comparison of the best ANN model with hybrid WANN model is illustrated in Figure3.

Through comparison, it's clearly that WANN has the most correlation to observed value compared with ANN model.

This is due to fact that wavelet decomposition reduces the complexity of streamflow time series. The findings related to this study, WANN has the strong ability to capture the variations of streamflow time series using previous values of streamflow combination with a high performance with ANN model.



. Figure 3. Comparison of WANN and ANN performance versus observed streamflow in the testing period

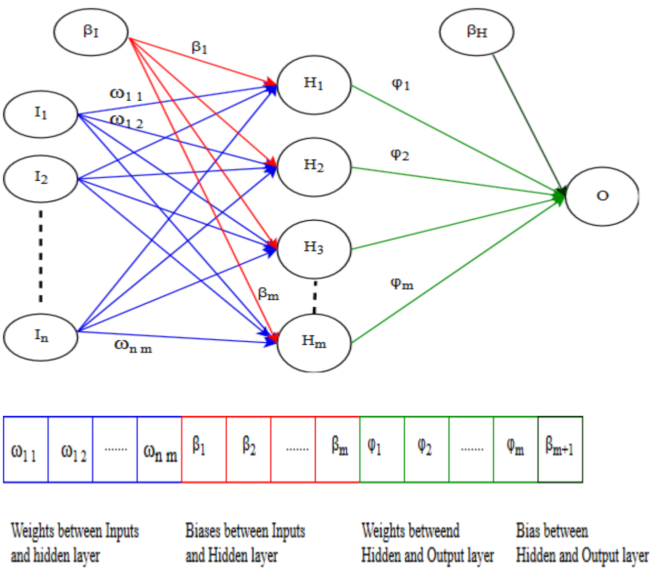


Figure 4. Simple ANN architecture

IV. CONCLUSION

In this paper, to predict one-month ahead forecast in the Bouchegouf gauge station in Seybouse watershed eastern part of Algeria, the wavelet transform was used to decompose monthly streamflow data into approximations and details, which were used as inputs to (ANN) model for developing hybrid (WANN) and compare it with the single (ANN).

According to the statistical performance criteria abovementioned, It is found that the conjunction of (WANN) is better and more efficient than the single (ANN) in all developed models with different input combinations,

This means that the performance increased after using discrete wavelet transform as a pre-processing tool. Moreover, a best accuracy has been achieved without increasing the number of inputs.

This new hybrid model presents an interesting prospect for streamflow forecasting and it can be a useful tool for hydrological and engineering prediction issues. In future works, some other input variables of climate such as evapotranspiration and precipitation can be used in forecasting of streamflow values and may increase the performance of the developed models [15].

REFERENCES

[1] A. Hossain et M. Nasser. « Comparison of the finite mixture of ARMA-GARCH. back propagation neural networks and support-vector machines in forecasting financial returns ». Journal of Applied Statistics. vol. 38. nº 3. p. 533‑ 551. mars 2011.

[2] M. Shafaei et O. Kisi. « Lake Level Forecasting Using Wavelet-SVR. Wavelet-ANFIS and Wavelet-ARMA Conjunction Models ». Water Resources Management. vol. 30. nº 1. p. 79‑ 97. janv. 2016.

[3] D.-I. Jeong et Y.-O. Kim. « Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction ». Hydrological Processes. vol. 19. nº 19. p. 3819‑ 3835. déc. 2005.

[4] Z. M. Yaseen. A. El-shafie. O. Jaafar. H. A. Afan. et K. N. Sayl. « Artificial intelligence based models for stream-flow forecasting: 2000–2015 ». Journal of Hydrology. vol. 530. p. 829‑ 844. nov. 2015.

[5] O. Kisi et M. Cimen. « A wavelet-support vector machine conjunction model for monthly streamflow forecasting ». Journal of Hydrology. vol. 399. nº 1‑ 2. p. 132‑ 140. mars 2011.

[6] A. Belayneh. J. Adamowski. B. Khalil. et B. Ozga-Zielinski. « Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models ». Journal of Hydrology. vol. 508. p. 418‑ 429. janv. 2014.

[7] S. Djerbouai et D. Souag-Gamane. « Drought Forecasting Using Neural Networks. Wavelet Neural Networks. and Stochastic Models: Case of the Algerois Basin in North Algeria ». Water Resources Management. vol. 30. nº 7. p. 2445‑ 2464. mai 2016.

[8] A. Danandeh Mehr. E. Kahya. A. Şahin. et M. J. Nazemosadat. « Successive-station monthly streamflow prediction using different artificial neural network algorithms ». International Journal of Environmental Science and Technology. vol. 12. nº 7. p. 2191‑ 2200. juill. 2015.

[9] Haykin. « Neural Networks: A Comprehensive Foundation ». Prentice Hall PTR Upper Saddle River. p. NJ. USA. 1994.

[10] Z. et al He. « A comparative study of artificial neural network. adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region ». Journal of Hydrology. vol. 509. p. 379‑ 386. févr. 2014.

[11] A. Altunkaynak et M. Ozger. « Comparison of Discrete and Continuous Wavelet–Multilayer Perceptron Methods for Daily Precipitation Prediction ». Journal of Hydrologic Engineering. vol. 21. nº 7. p. 04016014. juill. 2016.

[12] S. G. Mallat. « A theory for multiresolution signal decomposition: the wavelet representation ». IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 11. nº 7. p. 674‑ 693. juill. 1989.
[13] Ö. Kişi. « Stream flow forecasting using neuro-wavelet technique ». Hydrol. Process.. vol. 22. nº 20. p. 4142‑ 4152. sept. 2008.
[14] V. Nourani. M. Komasi. et A. Mano. « A Multivariate ANN-Wavelet Approach for Rainfall–Runoff Modeling ». Water Resources Management. vol. 23. nº 14. p. 2877‑ 2894. nov. 2009.
[15] R. M. Adnan, X. Yuan, O. Kisi, M. Adnan, et A. Mehmood, « Stream Flow Forecasting of Poorly Gauged Mountainous Watershed by Least Square Support Vector Machine, Fuzzy Genetic Algorithm and M5 Model Tree Using Climatic Data from Nearby Station », *Water Resources Management*, vol. 32, nº 14, p. 4469‑ 4486, nov. 2018.