

## PAPER 13 – Application of Smoothing Techniques in Stream Flow Prediction of Semi-Arid Region: A Case of Hadejia River, Nigeria

N.I. Abdullahi\*, S.J. Ijimdia, A.D. Adamu, and M.M. Muhammad

Department of Water Resources and Environmental Engineering; Ahmadu Bello University Zaria, Nigeria

\*Email: [nurafbg@gmail.com](mailto:nurafbg@gmail.com)

### ABSTRACT

This paper assessed the prediction capability of time series smoothing techniques in predicting the stream flow of Hadejia River. Four smoothing techniques were employed for the analysis of 36 years monthly stream flow data (1965 – 2000). The techniques are Moving Average (MAS), Holt and Winters (HWS), Simple Exponential (SES), and Fourier (FS) methods. The performance evaluation parameters used in assessing the suitability of the techniques were mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE), coefficient of determination ( $R^2$ ), and Nash-Sutcliffe Coefficient of Efficiency (NSE). The predicting capability of stream flow using MAS, HWS, and SES was found to be inadequate (MSE, RMSE, MAPE,  $R^2$ , and NSE values are out of range) whereas the FS technique was found to be suitable. Aside the month of June where  $R^2$  and NSE were 0.43, all other model's parameters of FS were adequate. The MSE was found to vary between 3.92 and 247.10, RMSE between 1.98 and 15.72, MAPE between 13.51 and 546.06,  $R^2$  between 0.434 and 0.923, and NSE between 0.430 and 0.923. The study concludes that Fourier smoothing technique is the most suitable for predicting the stream flow of the study area among all the techniques assessed.

**KEYWORDS:** Nash-Sutcliffe, River, Smoothing, Stream flow, and Root Mean Squared Error.

### 1 INTRODUCTION

Stream flow, or discharge, is the volume of water that moves over a designated point over a fixed period of time which is directly correlated to the amount of water moving off the watershed into the stream channel (EPA, 2012). Precise prediction of stream flow is of great importance for the effective management of water resources systems (Karran *et al.*, 2014; Adhikary *et al.*, 2018; and Güneş *et al.*, 2021). Among other benefits, stream flow prediction can influence important decisions such as water allocation, cropping strategies, water market planning, environmental watering, operating multi-storage water supply schemes, and drought management.

Many techniques have been applied in predicting stream flow values. Examples can be found in Post and Jakeman (1999) and Chiew *et al.* (2018) where they applied the Rainfall-Runoff model; Patel *et al.* (2016) and Mehraein *et al.*, (2022) applied different regression techniques; Adnan *et al.* (2017a); Liu *et al.* (2020) applied Artificial Neural Network; Parisouj *et al.*, (2022) applied coupled model by combining model-driven models and machine learning. Further examples could be found in Adnan *et al.*, (2017b); Machekposhti *et al.*, (2018) where they applied time series Autoregressive Integrated Moving Average (ARIMA) and Autoregressive Moving Average Model (ARMA) models. Time series smoothing models were also applied at various times. Examples can be found in Lukman and Tanan (2020), Papacharalampous and Tyralis (2020) and Brito *et al.* (2021). This paper is therefore set to assess the prediction capability of time series smoothing techniques in predicting the stream flow of Hadejia River, Nigeria.

### 2 THEORETICAL ANALYSIS

The smoothing technique is defined by an expression  $Y_t$  where  $t = 1, 2, \dots, n$  and the time series of concern defined by  $P_t Y_{t+h}$  where the predictor  $Y_{t+h}$  is that which returns minimum mean square error and  $\varepsilon_t$  that follows  $N(0,1)$  a white noise. Four smoothing methods are deployed in predicting the stream flow of the study area. The methods are; moving average, Holt and winters, simple exponential, and Fourier. The smoothing models are observed to be non-stationary that do not assume homoscedasticity (Rosenblad, 2021); the techniques chosen were thus all applied to the untransformed (raw) data.

#### 2.1 Moving Average Method

The moving average model takes into account past (and optionally future) observations to predict values and it works as a filter that removes white noise. An observation influences all future predictions and it is defined by Equation (2.1). The analysis utilized the equal weights method at 95% confidence level with  $q=2$

$$\begin{cases} Y_t = \mu_t + \varepsilon_t \\ \mu_t = \frac{\sum_{i=q}^{q+l} \omega_i Y_{t+i}}{\sum_{i=q}^{q+l} \omega_i} \end{cases} \quad 2.1$$

## 2.2 Holt's and winters (Linear Holt)

The linear Holt's model (sometimes referred to as Holt's-Winters non-seasonal algorithm) takes into account a trend and a permanent component that varies with time. The model is defined by Equation (2.2). The analysis was performed at 95% confidence level and  $\alpha$  and  $\beta = 0.2$ .

$$\begin{cases} Y_t = \mu_t + \beta_t t + \varepsilon_t \\ P_t Y_{t+h} = \mu_t + \beta_t t & h = 1, 2, \dots \\ S_t = \alpha Y_t + (1-\alpha)(S_{t-1} + T_{t-1}) & 0 < \alpha < 2 \\ T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1} & 0 < \beta < \frac{4}{\alpha-2} \\ Y_{t+h} = P_t Y_{t+h} = S_t + hT_t & h = 1, 2, \dots \end{cases} \quad 2.2$$

## 2.3 Simple Exponential

The exponential smoothing model takes into account past observations to predict values. The weight of the past observations decreases exponentially. The simple exponential model is defined by Equation (2.3). This study is performed at 95% confidence level with  $\alpha = 0.2$ .

$$\begin{cases} Y_t = \mu_t + \varepsilon_t \\ P_t Y_{t+h} = \mu_t & h = 1, 2, \dots \\ S_t Y_{t+h} = \alpha Y_t + (1-\alpha)S_{t-1} & 0 < \alpha < 2 \\ Y_{t+h} = P_t Y_{t+h} = S_t & h = 1, 2, \dots \end{cases} \quad 2.3$$

## 2.4 The Fourier Smoothing

The concept of the Fourier smoothing is to transform a time series into its Fourier coordinates, then remove part of the higher frequencies, and then transform the coordinates back to a signal. This new signal is a smoothed series.

## 3 MATERIALS AND METHODS

The paper utilized xlstat (an Excel extension) for the analysis. Four smoothing techniques were employed for the analysis of 36 years monthly stream flow data (1965 – 2000) of the Hadejia River to ascertain their suitability in predicting the flow of the river. The techniques are Moving Average (MAS), Holt and Winters (HWS), Simple Exponential (SES), and Fourier (FS) methods. The fitness of each method is ascertained by sets of performance evaluation parameters. The parameters were mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE), coefficient of determination ( $R^2$ ), and Nash-Sutcliffe Coefficient of Efficiency (NSE).

- i. The MSE measures the amount of error in statistical models. It assesses the average squared difference between the observed and [predicted values](#). When a model has no error, the MSE equals zero. As model error increases, its value increases. Values closer to zero represent a relatively fit model.
- ii. The RMSE is commonly used in model performance evaluation that computes and reported in the same unit as the model output of concern. It works well for continuous long-term simulations. The values of RMSE range between zero and infinity ( $0 - \infty$ ) with 0 as the optimal value (Moriassi *et al.*, 2015). Values closer to zero represent a relatively fitter model.
- iii. The MAPE measures the accuracy of a forecast system. It measures this [accuracy](#) as a [percentage](#) and can be calculated as the average absolute per cent error for each time period minus actual values divided by actual values. MAPE values range from zero to infinity ( $0 - \infty$ ), where the lower the value the more accurate the prediction is.
- iv.  $R^2$  is widely used in hydrological modelling studies, thus serving as a benchmark for performance evaluation. The values of  $R^2$  range between zero and one ( $0 - 1$ ). For a monthly flow simulation,  $R^2 >$

0.85 shows a very good model,  $0.75 < R^2 \leq 0.85$  a good model,  $0.60 < R^2 \leq 0.75$  a satisfactory model, and  $R^2 \leq 0.60$  shows not satisfactory model (Moriassi *et al.*, 2015).

- v. The NSE is used to evaluate model performance for several output responses (including stream flow) and can also be used to determine how well a model simulates trends for output responses of concern. The values of NSE range between minus infinity and one ( $-\infty - 1.0$ ) with an optimal value of 1.0. For a monthly flow simulation,  $NSE > 0.8$  shows a very good model,  $0.70 < NSE \leq 0.8$  shows a good model,  $0.50 < NSE \leq 0.70$  a satisfactory model, and  $NSE \leq 0.5$  shows not satisfactory model (Moriassi *et al.*, 2015).

## 4 RESULTS AND DISCUSSION

The stream flow of the Hadejia River (1965 – 2000) at Hadejia Bridge Station is predicted using four time series smoothing modelling techniques; MAS, SES, HWS, and FS. The Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Coefficient of Determination, and Nash-Sutcliffe Coefficient of Efficiency (NSE) were the parameters used to assess the performance of the models.

### 4.1 Moving Average (MAS)

The stream flow prediction using MAS returns a result that can best be described as an unfit model. As presented in Table 4.1, the model initially shows some glimpse of fitness with  $R^2$  values in the first three months (January, February, and March) and returns a good fit model with values of 0.831, 0.813, and 0.759 respectively and the last two months (November and December) with values of 0.880 and 0.847 respectively. The MSE, RMSE, and NSE in the first month also return promising values at 9, 2.93, 87.93 and 0.827 respectively. Only January has a satisfactory NSE value. The MSE, RMSE, and MAPE values also return values that are not close to zero. It could be seen from the performance evaluation parameters that the model was only able to predict the flow in January, a period of low flow. The poor performance of the model could be as a result of the disadvantage of the technique in applying it to a data series that have trend and seasonality components and its suitability for relatively short-term forecasting as reported by Lukman and Tanan (2020).

### 4.2 Holt's and Winters (HWS)

The HWS shows a relatively better prediction capability than MAS. As presented in Table 4.1, aside  $R^2$  which varies between 0.030 and 0.481 all other performance evaluation parameters return relatively better results than that of MAS. The MSE results vary between 28.75 and 687, RMSE between 4.564 and 26.223, MAPE between 32.024 and 355.772, and NSE between -0.451 and 0.481. The HWS model also predicts better during the period of low flows as results of  $R^2$ , NSE, RMSE, and MSE are better from January to April (even though the  $R^2$  and NSE were not adequate) which corresponds to the period of low flow. The result agrees with the work of Toro *et al.*, (2008) where HWS performed relatively better than MAS and lower than SES (as is also the case here which can be understood better after reading the next paragraph).

### 4.3 Simple Exponential Smoothing (SES)

The SES shows a relatively better prediction capability than HWS. As presented in Table 4.2, the MSE varies between 17.853 and 721.072, RMSE between 4.225 and 26.853, MAPE between 32.365 and 333.473,  $R^2$  between 0.005 and 0.621, and NSE between -0.064 and 0.415. The model also predicts better during a period of low flow as the performance evaluation parameters return a better result between January and April. The result is in agreement with the work of Lukman and Tanan (2020) where SES was observed to be weak in predicting monthly stream flow by returning the same average discharge value for each month (January – December). In the work, the model accuracy was found to be -0.08 while the MAPE value was found to be 142.55 which are all within the range of the values found by this study. The result is also in agreement with Toro *et al.*, (2023) where SES predicts stream flow relatively better than HWS. Overall, the model performance is also inadequate in predicting the stream flow of the study area. For this limitation, Dullah *et al.*, (2023) proposed the integration of a nonlinear autoregressive neural network with HWS and SES which improves the precision of the prediction beyond applying the HWS or SES alone. Overall, the predicting capability of stream flow of the study area using HWS is inadequate.

### 4.4 Fourier (FS)

The FS technique returns a much better predicting capability in the study area than the previous three models (MAS, HWS, and SES). Aside the month of June where  $R^2$  and NSE were 0.43, all other model parameters are

good. The MSE was found to vary between 3.92 and 247.10, RMSE between 1.98 and 15.72, MAPE between 13.51 and 546.06,  $R^2$  between 0.434 and 0.923, and NSE between 0.430 and 0.923. The model also shows greater prediction capabilities in the period of low flow even though its predicting capabilities for the remaining months (except June) are adequate. The model's performance evaluation parameters are presented in Table 4.2 while Fig. 4.1 presents the results showing the observed versus simulated stream flow across all months. This result can be supported by the work of Bayazit et al., (2001) where they found Fourier analysis to be able to simulate monthly stream flow relatively well with reasonably small variations within the simulations.

## 5 CONCLUSIONS

This paper assessed the prediction capability of four time series smoothing techniques (moving average, Holt and winters (linear), simple exponential, and Fourier) in predicting a monthly stream flow of Hadejia River, Nigeria. The predicting capabilities of moving average, Holt and winters, and simple exponential were found to be inadequate. The Fourier method returns a relatively better technique for predicting the stream flow of the river. The Fourier performance evaluation parameters return RMSE value in the range of 1.98 and 13.51,  $R^2$  and NSE values in the range of 0.4 and 0.9. This paper, therefore, recommends the application of Fourier in predicting of monthly stream flow of the Hadejia River

Table 4.1: Performance Evaluation Parameters of MAS and HWS

	Moving Average (Equal Weight Method)					Holt and winters (Linear Holt)				
	MSE	RMSE	MAPE	R <sup>2</sup>	NSE	MSE	RMSE	MAPE	R <sup>2</sup>	NSE
Jan	9	2.93	87.25	0.8310	0.827	28.754	5.362	95.254	0.544	0.481
Feb	98	9.88	197.35	0.8130	-1.661	24.096	4.909	78.273	0.433	0.406
Mar	453	21.29	485.61	0.7590	-9.134	30.413	5.515	84.124	0.418	0.397
Apr	1529	39.11	1033.31	0.4570	-54.176	20.834	4.564	129.466	0.354	0.332
May	4091	63.96	753.90	0.3820	-107.96	35.882	5.990	244.330	0.176	0.151
Jun	11612	107.76	568.83	0.0002	-113.525	150.934	12.286	355.772	0.030	-0.387
July	32130	179.25	303.59	0.0170	-0.217	150.849	12.282	36.356	0.050	-0.301
Aug	91420	302.36	266.30	0.0003	-308.587	291.167	17.064	32.024	0.222	-0.399
Sep	262559	512.40	295.91	0.2390	-359.085	681.905	26.113	36.995	0.268	-0.415
Oct	779165	882.70	1220.54	0.4590	-1011.316	687.656	26.223	74.955	0.235	0.038
Nov	2196079	1481.92	5455.70	0.8800	-7891.579	253.197	15.912	88.477	0.388	0.181
Dec	6099860	2469.79	21138.40	0.8470	-49721.141	105.049	10.249	186.070	0.366	0.236

Table 4.2: Performance Evaluation Parameters of SES and FS

	Simple Exponential					FOURIER				
	MSE	RMSE	MAPE	R <sup>2</sup>	NSE	MSE	RMSE	MAPE	R <sup>2</sup>	NSE
Jan	29.950	5.473	96.510	0.621	0.415	3.92	1.98	70.50	0.923	0.923
Feb	23.137	4.810	62.855	0.532	0.387	3.63	1.90	176.59	0.904	0.904
Mar	29.857	5.464	58.847	0.491	0.351	11.53	3.40	546.06	0.749	0.749
Apr	17.853	4.225	90.635	0.454	0.374	4.99	2.23	176.01	0.825	0.825
May	31.426	5.606	182.739	0.260	0.186	13.99	3.74	105.25	0.638	0.638
Jun	119.122	10.914	333.473	0.005	-0.201	59.07	7.69	170.68	0.434	0.430
July	150.195	12.255	37.736	0.014	-0.221	48.81	6.99	21.01	0.677	0.674
Aug	250.000	15.811	32.365	0.058	-0.345	57.83	7.60	13.51	0.810	0.810
Sep	644.468	25.386	40.994	0.054	-0.227	247.10	15.72	25.22	0.671	0.671
Oct	721.072	26.853	111.252	0.052	-0.064	218.71	14.79	64.09	0.724	0.724
Nov	250.598	15.830	91.057	0.335	0.115	90.84	9.53	84.10	0.683	0.683
Dec	103.778	10.187	183.425	0.404	0.175	46.11	6.79	148.04	0.635	0.635

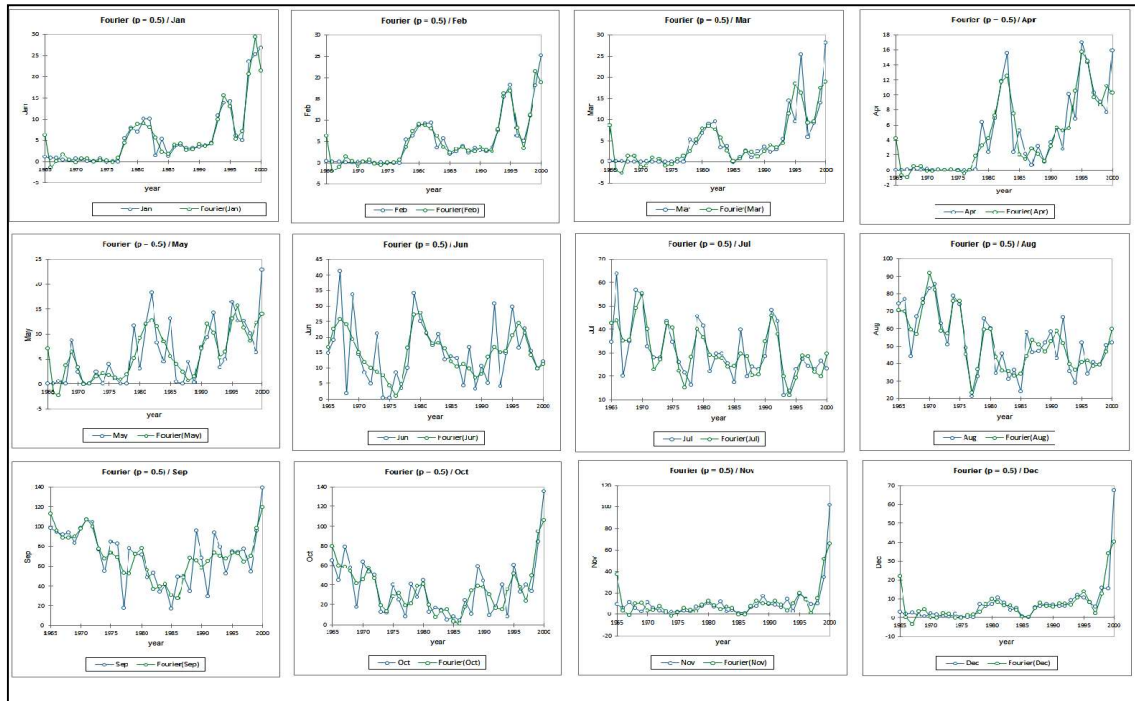


Fig 4.1: Observed Versus Simulated Stream Flow Across All Months.

## REFERENCES

- Adhikary, S. K., Muttill, N., & Yilmaz, A. G. (2018). Improving streamflow forecast using optimal rain gauge network-based input to artificial neural network models. *Hydrology Research*, 49(5), 1559–1577. <https://doi.org/10.2166/nh.2017.108>
- Adnan, R. M., Yuan, X., Kisi, O., & Yuan, Y. (2017a). Streamflow Forecasting Using Artificial Neural Network and Support Vector Machine Models. *American Scientific Research Journal for Engineering*. <http://asrjetsjournal.org/>
- Adnan, R. M., Yuan, X., Kisi, O., & Curtef, V. (2017b). Application of Time Series Models for Streamflow Forecasting. *Civil and Environmental Research*, 9(3), 56–63. [www.iiste.org](http://www.iiste.org)
- Alonso Brito, G. R., Rivero Villaverde, A., Lau Quan, A., & Ruiz Pérez, M. E. (2021). Comparison between SARIMA and Holt–Winters models for forecasting monthly streamflow in the western region of Cuba. *SN Applied Sciences*, 3(6). <https://doi.org/10.1007/s42452-021-04667-5>
- Bayazit, M., ÖNÖZ, B., & Aksoy, H. (2001). Nonparametric stream flow simulation by wavelet or Fourier analysis. *Hydrological Sciences Journal*, 46(4), 623–634. <https://doi.org/10.1080/02626660109492855>
- Chiew, F. H. S., Zheng, H., & Potter, N. J. (2018). Rainfall-Runoff modelling considerations to predict streamflow characteristics in ungauged catchments and under climate change. *Water (Switzerland)*, 10(10). <https://doi.org/10.3390/w10101319>
- Dullah, H., Ahmed, A.N., Kumar, P. et al. (2023). Integrated nonlinear autoregressive neural network and Holt winters exponential smoothing for river streaming flow forecasting at Aswan High. *Earth Sci Inform* 16, 773–786. <https://doi.org/10.1007/s12145-022-00913-5>
- EPA (2012). *Stream Flow: Monitoring & Assessment*. United State Environmental Protection Agency. <https://archive.epa.gov/water/archive/web/html/vms51.html>
- Güneş, M. Ş., Parim, C., Yıldız, D., & Büyüklü, A. H. (2021). Predicting monthly streamflow using a hybrid wavelet neural network: Case study of the Çoruh river basin. *Polish Journal of Environmental Studies*, 30(4), 3065–3075. <https://doi.org/10.15244/pjoes/130767>

- Karran, D. J., Morin, E., & Adamowski, J. (2014). Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes. *Journal of Hydroinformatics*, 16(3), 671–689. <https://doi.org/10.2166/hydro.2013.042>
- Kumar, M., Jawaharlal, H., Krishi Vishwavidyalaya, N., Seetpal, M., Patel, S., Hardaha, M. K., Seetpal, M. K., & Madankar, K. K. (2016). Multiple Linear Regression Model for Stream Flow Estimation of Wainganga River. *American Journal of Water Science and Engineering*, 2(1), 1–5. <https://doi.org/10.11648/j.ajwse.20160201.11>
- Liu, D., Jiang, W., Mu, L., & Wang, S. (2020). Streamflow Prediction Using Deep Learning Neural Network: Case Study of Yangtze River. *IEEE Access*, 8, 90069–90086. <https://doi.org/10.1109/ACCESS.2020.2993874>
- Lukman, M., & Tanan, B. (2021). Time Series Modeling by using Exponential Smoothing Technique For River Flow Discharge Forecasting : A Case Study of Cabenge, Walanae, and Cenranae rivers system. *IOP Conference Series: Materials Science and Engineering*, 1088(1), 012100. <https://doi.org/10.1088/1757-899x/1088/1/012100>
- Machekposhti, K. M., Sedghi, H., Telvari, A., & Babazadeh, H. (2018). Flood Predicting in Karkheh River Basin Using Stochastic ARIMA Model. *International Journal of Agricultural and Biosystems Engineering*, 12(3), 84–91.
- Mehraein, M., Mohanavelu, A., Naganna, S. R., Kulls, C., & Kisi, O. (2022). Monthly Streamflow Prediction by Metaheuristic Regression Approaches Considering Satellite Precipitation Data. *Water (Switzerland)*, 14(22). <https://doi.org/10.3390/w14223636>
- Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, 58(6), 1763–1785. <https://doi.org/10.13031/trans.58.10715>
- Papacharalampous, G., & Tyralis, H. (2020). Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability. *Journal of Hydrology*, 590. <https://doi.org/10.1016/j.jhydrol.2020.125205>
- Parisouj, P., Mokari, E., Mohebzadeh, H., Goharnejad, H., Jun, C., Oh, J., & Bateni, S. M. (2022). Physics-Informed Data-Driven Model for Predicting Streamflow: A Case Study of the Voshmgir Basin, Iran. *Applied Sciences (Switzerland)*, 12(15). <https://doi.org/10.3390/app12157464>
- Post, D. A., & Jakeman, A. J. (1999). Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modelling*, 123, 91–104. [www.elsevier.com/locate/ecolmodel](http://www.elsevier.com/locate/ecolmodel)
- Rosenblad, A.K. (2021): Accuracy of automatic forecasting methods for univariate time series data: A case study predicting the results of the 2018 Swedish general election using decades-long data series. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 7:3, 475-493, DOI: 10.1080/23737484.2021.1964407
- Toro, F. C. H., González, D., Benedicto, P., González, S., Fernández, F., & Riverola. (2008). Water flows modeling and forecasting using a RBF neural network. *Julio*, 6(12), 13–31.