Journal of Environmental Treatment Techniques 2021, Volume 9, Issue 4, Pages: 796-803

Journal of ENVIRONMENTAL Treatment Techniques

J. Environ. Treat. Tech.
ISSN: 2309-1185

Journal web link: https://dormaj.org/index.php/jett
https://doi.org/10.47277/JETT/9(4)803

# Development of Regression-based Hydrologic Model for Estimating Inflows to Tarbela Reservoir

**Noor Yaseen [1*], Habib-Ur-Rehman [2], Salik Haroon Abbasi [3]**

[1] Graduate Assistant; M.Sc. Student, Civil Engineering Department, University of Engineering and Technology, Lahore, Pakistan
[2] Professor & Dean of Civil Engineering Department, University of Engineering and Technology, Lahore, Pakistan
[3] Graduate Assistant; M.Sc. Student, University of Engineering and Technology, Lahore, Pakistan

Received: 19/12/2020 Accepted: 22/10/2021 Published: 20/12/2021

## Abstract

The assessment of daily discharges at the Tarbela Dam is one of the major concerns for the reservoir operational team. As per record, the existing model used by Tarbela Dam Project (TDP) Engineers has not been reliably estimating the inflows to the reservoir as compared to those obtained from the reservoir operational data. This paper explores the development of a new hydrologic model using regression techniques. For this, four years of representative data for the period from 2013 to 2016 were obtained for daily inflows at four gauging stations namely, Indus river at Tarbela, Indus river at Besham Qila, Siran river near Phulra, and Brandu iver near Daggar. Both multiple linear regression (MLR) and multiple nonlinear regression (MNLR) were performed to develop models taking inflows at Tarbela as the response variable and inflows at the remaining three upstream-gauging stations as the explanatory variables. Based on several statistical measures and the visual inspection of the testing models, the MNLR provided a better representation of the relationship between the Tarbela inflows and the upstream-gauging stations' inflows. The best-fit nonlinear model declared the inflows at Besham as the most influential explanatory variable followed by the inflows at Phulra, while eliminating those at Daggar, suggesting that the inflows to Tarbela can effectively be estimated without the inclusion of Daggar inflows. The outcomes of the newly developed nonlinear model are considerably better in comparison to those of the existing model used by TDP Engineers. This study is helpful for the reservoir operational team to estimate the daily flows based on upstream-gauging stations data; it is recommended to update the model to estimate inflows to the reservoir for every three to four years.

**Keywords**: Hydrologic Model; Tarbela Reservoir; Upstream-gauging Stations; Multiple Linear Regression; Multiple Nonlinear Regression

## 1 Introduction

Developing a hydrological model for a reservoir plays an important role in finding out the solutions toward problems related to water and environmental resource management such as forecasting of hydrological events [1] and groundwater level forecasting [2]. Although various new methods are used for hydrologic forecasting purposes such as the wavelet transforms [3] and hybrid wavelet-artificial intelligence models [4], the regression technique is still considered as one the most widely used yet simple and reliable methods for hydrological modeling. Also, the demand for using inexpensive models is greatly evolved over the last decade to come up with simple models requiring a limited number of parameters [5]. For example, with the help of only upstream data, Pizarro-Tapia et al. [6] calibrated and validated the linear and nonlinear models to estimate downstream flooding in two Mediterranean river basins in Chile.

Like other developing countries in the Asia-Pacific region, Pakistan is facing water crises. As Pakistan is an agricultural country and therefore for better management of water, proper strategies should be made that encompass the prediction of discharges for important hydraulic structures such as the Tarbela reservoir. Tarbela Dam is being maintained and operated by Tarbela Dam Project (TDP), whereas distributaries of the Indus River are operated by the Indus River System Authority (IRSA). The inflow coming to Tarbela is measured at three tributaries: Indus River at Besham Qila, Siran river near Phulra, and Brandu river near Daggar (see Fig. 1). Among these three, the major contribution in the inflows to Tarbela is that of Besham Qila—a gauging station about 65 km above the top end of the reservoir—which is about 93% of the Indus River flow to the reservoir [7]. The inflow at Tarbela reservoir is measured through the use of a simple 'storage change and outflow' equation on daily basis, given in Eq. (1):

$$T = \frac{\partial S}{\partial t} + O \qquad (1)$$

where $T$ is the inflow at Tarbela from the reservoir data, $\partial S/\partial t$ is the storage change of the reservoir per day and is obtained from the elevation area capacity curves or tables and $O$ represents the total daily outflows measured at the outlets of the dam. Also, the TDP Engineers estimate the inflow at Tarbela from three upstream-gauging stations data as illustrated in [8]; their equation is termed herein as the existing model, given in Eq. (2):

$$T = B + 2(P + D) \qquad (2)$$

where $T$ is the Tarbela inflow, $B$ is the Indus river discharge at Besham Qila lagged by 7 hours, $P$ is the Siran river discharge near Phulra lagged by 6 hours and $D$ is the Brandu river

**Corresponding author:** Noor Yaseen, Graduate Assistant; M.Sc. Student, Civil Engineering Department, University of Engineering and Technology, Lahore, Pakistan, Email: noor.yaseen@uet.edu.pk

discharge near Daggar lagged by 6 hours. The representatives of TDP are bound to release water following the indents placed by IRSA, as the irrigation water has priority over other uses (including the power generation) as per Water Apportionment Accord (WAA) 1991 [9]. Eq. (2) as per record is not reliable and truly applicable as it undergoes great temporal variation between the estimated and the actual-observed discharges. This was the main complaint of IRSA that the inflows estimated by TDP Engineers (Eq. (2)) do not match well with those calculated from outflows and storage change in the reservoir (Eq. (1)) [10]. Similarly, a recent study by Ullah et al [11] conducted an artificial neural network (ANN) on seasonal flow data from the year 2015 to 2019 for forecasting water discharges for the Tarbela dam. However, these authors neither presented the simulated flows with the observed ones nor presented the models in mathematical or statistical form. Moreover, the equations obtained through the ANN method are susceptible to the over-fitting problem due to their complex network structure [12]. Hence, it is necessary to develop a regression-based model to yield reliable and accurate results for estimating inflows at Tarbela based on recently recorded upstream-gauging stations data.

The existing literature review proposes that hydrological modeling be investigated in such a way as to obtain results closest to the observed data and that the model should remain inexpensive in terms of parameters i.e., there should be the least model complexity. The development of such models is possible with the help of regression analysis. To the best of our knowledge, very few studies are available for the case, where hydrologic models are developed for the prediction of inflows to Tarbela reservoir such as the HBV model application to the upper basins of Indus river for forecasting inflows to Tarbela dam [13]. However, these types of models need several input parameters such as precipitation observations, temperature data, and long-term estimates of potential evapotranspiration, which increase the complexity of the model. Also, such models are useless in case of the unavailability of the precipitation and temperature data as a result of the poor working conditions of the meteorological stations. In addition, the Tarbela reservoir itself is facing significant issues among which the delta deposits is of prime concern since the reservoir has already lost its storage capacity up to 40.58% due to the high influx of the sediments from the Upper Indus Baisn [14]. On the other hand, the demand for water for either electricity generation or irrigation purposes is increasing on daily basis: a recent paper by Amin et al [15] revealed that the future water demand in the Upper Indus Basin by 2050, is predicted to be more than double as compared with that in 2015. To cope with such aforementioned issues of Tarbela reservoir, the estimation of inflows to Tarbela has investigated herein the current research work, based on upstream-gauging stations data comprising discharge values only. This was achieved by applying both Multiple Linear Regression (MLR) and Multiple Nonlinear regression (MNLR) to model inflows to Tarbela for the gauging station's data of 4 years (2013-2016). Different statistical measures were computed for assessing the model performance of the newly developed regression-based equations. Finally, the simulated and observed inflows to the reservoir were visually inspected through scatter plots and time series plots.

## 2 Study Area

Tarbela Dam is an earth- and rock-fill dam along the Indus river, situated mainly in Sawabi and partly in the Haripur district of Khyber Pukhtunkhwa, Pakistan with Latitude: 34◦05′23” N, Longitude: 72◦41′54” E. The main dam is about 105 km northwest of Islamabad having a length of 274 m and a height of 143 m [10]. The upstream of Tarbela reservoir constitutes an area of 169650 km$^2$, of which 90% lies between the Karakoram and Himalaya ranges [7, 16]. Due to the high elevation of the Indus basin, most of the precipitations occur in the form of ice and snow, which result in the formation of glaciers. The major part of the 80% inflow into the Indus river is due to the melting of these glaciers [9, 17]. According to reports, the average annual inflow to the Tarbela reservoir is 81 BCM [18]. This inflow is mainly contributed by three major tributaries: Indus river at Besham Qila, Siran river near Phulra, and Brandu river near Daggar. According to the report published by Surface Water Hydrology Project (SWHP), the drainage areas of the Indus river at Besham Qila, Siran river near Phulra, and Brandu river near Daggar are 162393 km$^2$, 1057 km$^2$, and 599 km$^2$, respectively [19].

Figure 1: Google map of Tarbela Reservoir

## 3 Methodology

To develop a new hydrologic model for the Tarbela reservoir, representative inflows at stream gauging stations of Besham Qila, Phulra, and Daggar, and inflows measured at Tarbela were collected from the Surface Water Hydrology Project (SWHP) of Water and Power Development Authority (WAPDA) for the four years from January 2013 to December 2016. After the collection of data, the first step was to computerize the data that was in the form of pictures. Minitab—a statistical tool for analyzing data—was used in the research work for easy retrieval and analysis of the data. Both correlation and regression were conducted to find out how the inflow at Tarbela would vary with the discharges at Besham Qila, Phulra, and Daggar.

The strength of the correlation between the response variable and the individual explanatory variables was assessed using the Pearson correlation coefficient (r), which is a correlation-based statistical measure. The correlation coefficient can be either negative or positive, and it varies from -1 to +1, respectively; the negative and positive signs demonstrate the corresponding negative linear correlation and positive linear correlation [20, 21]. The formula for calculating the correlation coefficient is given by Eq. (3).

$$r = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum(y_i - \bar{y})^2 \sum(x_i - \bar{x})^2}} \tag{3}$$

where, $y_i$ is the observed values of the response variable; $x_i$ is the observed values of the explanatory variable; $\bar{y}$ is the mean of observed $y_i$ values; and $\bar{x}$ is the mean of observed $x_i$ values. Once the correlation was conducted, it was followed by regression, in which case both Multiple Linear Regression (MLR) and Multiple Nonlinear Regression (MNLR) were performed to develop a hydrologic model for estimating inflows to Tarbela reservoir. To examine the overall significance of a linear model, the global test (F-test) was conducted, whereas, for the individual significance of the explanatory variables to be evaluated, the p-value approach and the t-distribution were conducted. To assess the significance of the parameters for the case of the nonlinear models, the 95% Confidence Interval (95% CI) approach was used. The models were compared based on two statistical measures, one of which is the absolute error measure i.e., the Root Mean Square Error (RMSE), and second is the relative error measure known as Nash-Sutcliffe Efficiency (NSE). The Root Mean Square Error, as given in Eq. (4), is a nonnegative statistic having no upper bound and is widely used to demonstrate how far the fitted values are from the observed values of the response variable [22]. The RMSE gives better decision in that it describes the model performance in terms of units of the variable. The smaller the value of RMSE, the closer the data points fall to the fitted curve. The Nash-Sutcliffe Efficiency is a normalized statistic (given in Eq. (5)) that indicates how well the simulated values fit the observed values [23]. In other words, it demonstrates how well the plot of simulated versus observed data fits the 1:1 line. The N-SE ranges from negative infinity to 1; the closer the value of N-SE to 1, the better is the model performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(T_{est} - T_{obs})^2}{N}} \tag{4}$$

$$\mathrm{N-SE} = 1 - \frac{\sum(T_{est} - \bar{T}_{obs})^2}{\sum(T_{obs} - \bar{T}_{obs})^2} \tag{5}$$

where, $T_{est}$ is the estimated or simulated discharge, $T_{obs}$ is the observed discharge, $\bar{T}_{obs}$ is the mean of observed discharge values, and N is the number of observed data entries. The best-fit model was selected as the one having the least RMSE and highest N-SE. Besides, scatter plots containing trend lines and 1:1 lines and the time-series graphs were used to assess the adequacy of the final-proposed model by comparing the simulated values with the observed data. The stepwise procedure for developing the regression-based hydrologic model is illustrated in Fig. 2.

## 4 Results and Discussion

### *4.1 Correlations*

In this section, the correlation was highlighted by assessing the association between the response and explanatory variables. As mentioned earlier in the Introduction Section that the major inflow to Tarbela is approaching through Besham Qila, and the same was observed by correlating the measured inflow at Tarbela (T) with that of each of the three upstream-gauging stations i.e., Besham Qila (B), Phulra (P), and Daggar (D) (see Fig. 3 and Table 1).

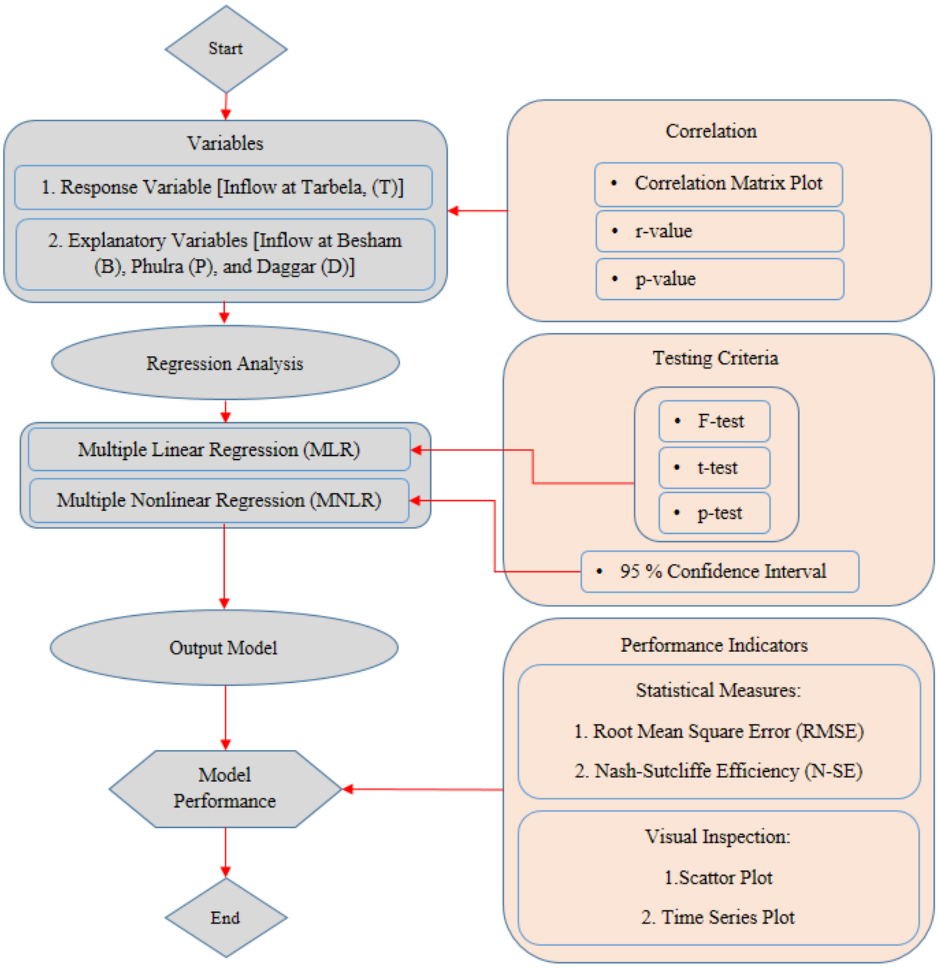

Figure 2: Methodology for development of regression-based hydrologic model

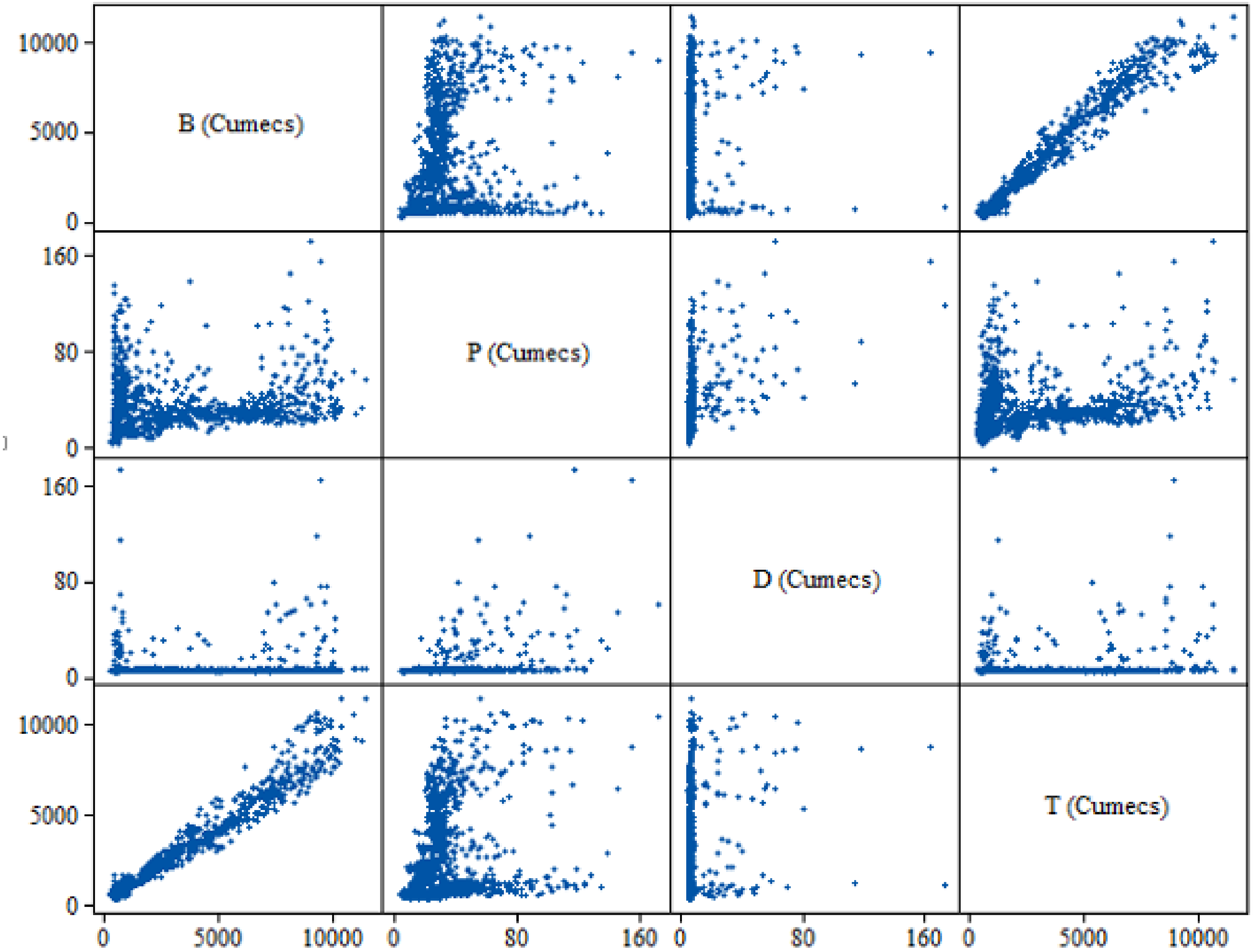


Figure 3: Scatter plot matrix of B, P, D, and T for assessing the strength of correlations

Table 1: Pearson correlation coefficient and p values: B, P, D, and T

| | B | P | D |
|---|---|---|---|
| P | 0.261 (0.000) | — | — |
| D | 0.160 (0.000) | 0.391 (0.000) | — |
| T | 0.982 (0.000) | 0.311 (0.000) | 0.179 (0.000) |

*Pearson correlation coefficient, r*
*Values in the brackets show the p-value*

From Fig. 3 and Table 1, it is clear that out of the three explanatory variables, B is the most influential with r of 98.20%, while D is the least influential explanatory variable with r of 16%. While P and D are not individually well-correlated with T, they can contribute mildly in fitting the model when T is regressed on them in combination with B. It is also obvious from either the bottom-left corner or the top-right corner diagram of Fig. 3 that the recorded discharges at Besham are greater than the recorded inflows to Tarbela (though Tarbela has a greater catchment). This is primarily due to the measurement errors at the gauging stations of Besham, as pointed out by [10, 13].

### *4.2 Multiple linear regression (MLR)*

Before developing a hydrologic model, the existing model given in Eq. (2) was analyzed; the average prediction error for this model is 266 Cumecs and this model overestimates the inflows by approximately 10%. The overestimation of inflows is obvious from Fig. 4 since the trend line is well above the 1:1 line. This model was also analyzed by the experts of the Sixth Periodic Inspection of the Tarbela dam [10], in which case the results of the model were not matching well with the observed inflows for the year 2014 to 2016. The existing model (Eq. (2)) exhibits an RMSE of 657 Cumecs and an N-SE of 93.48%. However, at this stage, it is difficult to determine how large or small these values are. Even by looking at Fig. 4, it is difficult to demonstrate whether the deviation between the trend line and the 1:1 line is higher. The RMSE and N-SE along with the visual inspection using trend line and 1:1 line will be compared with those of the other models that are subsequently developed in this research.

For model-building, a multiple linear regression was applied using the backward elimination method, meaning that the full model incorporating all the explanatory variables and the intercept term was established first, followed by successive deletion of the insignificant predictors one at a time, using t-statistics and p-statistics. The model parameters were estimated based on minimizing the sum of squared errors (SSE)—the procedure being known as the method of ordinary least squares. The newly developed model using multiple linear regression analysis is given in Eq. (6), which incorporates all the three explanatory variables along with an intercept:

$$T = -27 + 0.85937B + 6.368P + 0.38D \quad (6)$$

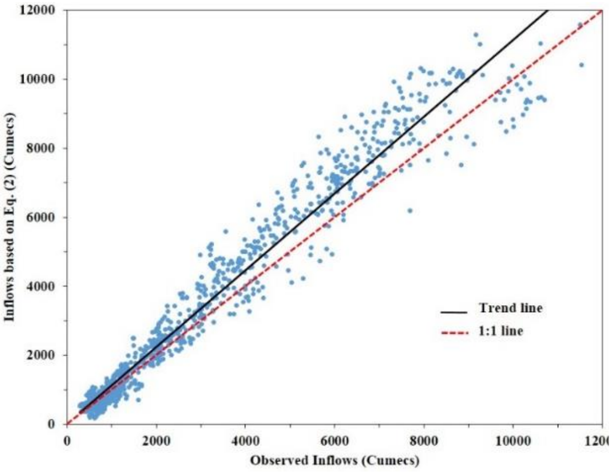


Figure 4: Scatter plot for comparing inflows from Eq. (2) and the observed flows

As for assessing the adequacy of the model, it can be observed (see Table 2) that the RMSE of 453.080 Cumecs has been significantly reduced as compared with that of the existing model (Eq. (2)) having an RMSE of 657 Cumecs. Also, the N-SE has been increased from 93.48% to 96.91%, indicating that the simulated values of Eq. (6) are fitting the observed values with comparatively higher accuracy. The overall significance of the model is checked based on the global F-test. Since the F-value of the regression model is 13743.25, which is much higher than the critical F-value (Fcv) of 2.60; consequently, the model is significant overall. (Note that this value is based on critical F-distribution and is taken from Table A.6 of Walpole et al. [24]; similarly, the critical values of the t-distribution (tcv) are taken from Table A.4 of the same reference.) As for the individual contribution of the explanatory variables, two criteria are applied: the p-value approach and the t-distribution test. The p-value for both the constant term and D is above the critical significance level of pcv = 0.05, which means that these two terms showed no statically significant relationship with the response variable, T. Before running regression analysis, the variable D was discarded because of its statistical insignificance in the model (p-value = 0.761 > pcv = 0.05 and t-value = 0.30 < tcv = 1.96). This may mean that changing D has very little to no effect on changes in Tarbela inflow. The same is obvious from Fig.1 and its much smaller r-value of 0.178 given in Table 1, confirming no linear relationship of it with T. Eliminating D will also help neglect the measurement of the least influential variable, D, suggesting (alternatively) that the concerned authority should give due importance to correctly measure the most influential variable, B. The resulting multiple linear models after excluding D are given in Eq. (7).

$$T = -26.5 + 0.85947B + 6.435P \quad (7)$$

From Table 3, it can be observed that there is no significant change in either the RMSE or N-SE as compared to those of Eq. (6). However, it is noticed that the constant term is statistically insignificant (p-value = 0.209 > 0.05 and t-value = -1.26 < 1.96). Hereafter, to reveal the effect of discarding the constant term on the model, the regression analysis was further performed without the incorporation of the intercept, and the following model resulted:

$$T = 0.85747B + 6.009P \quad (8)$$

By looking at Table 4, it is shown that by dropping D and the constant term, a best-fit model was yielded as both the explanatory variables (B and P) are statistically significant in that their p-values and t-values are less than 0.05 and greater than 1.96, respectively. However, eliminating the constant term (and merely D) could lead us to lose information that may help explain the variance of and forecast the magnitudes of the response variable. This is also noticeable by the RMSE and N-SE being not substantially changed as compared with the previous model i.e., Eq. (7). In sum, although Eq. (7) is equivalent to Eq. (6) and Eq. (8) in the viewpoint of almost equal N-SE and RMSE, it is the best fit model among all the linear models because it exhibited the least RMSE and highest N-SE value.

Table 2: Regression results of Eq. (6)

| **Term** | **Coefficient** | **t-value** | **p-value** |
|---|---|---|---|
| Constant | -27.0 | -1.27 | 0.203 |
| B | 0.85937 | 192.19 | 0.000 |
| P | 6.368 | 10.78 | 0.000 |
| D | 0.38 | 0.30 | 0.761 |
| S = 453.080 | E = 0.969 | F-value = 13743.25 | |

Fcv = 2.60, tcv = 1.96, pcv = 0.05

Table 3: Regression results of Eq. (7).

| **Term** | **Coefficient** | **t-value** | **p-value** |
|---|---|---|---|
| Constant | -26.5 | -1.26 | 0.209 |
| B | 0.85947 | 192.78 | 0.000 |
| P | 6.435 | 11.72 | 0.000 |
| S = 452.277 | E = 0.969 | F-value = 20651.64 | |

Fcv = 3.00, tcv = 1.96, pcv = 0.05

Table 4: Regression results of Eq. (8)

| **Term** | **Coefficient** | **t-value** | **p-value** |
|---|---|---|---|
| B | 0.85747 | 205.86 | 0.000 |
| P | 6.009 | 13.89 | 0.000 |
| S = 452.870 | E = 0.969 | F-value = 40475.51 | |

Fcv = 3.84, tcv = 1.96, pcv = 0.05

### *4.3 Multiple Nonlinear Regression (MNLR)*

It is well known that nonlinear models convey more information than do linear models. For the nonlinear regression analysis, instead of calculating the p-value and t-value for the individual parameter estimates, their significance was evaluated in terms of confidence intervals. And, for the overall significance of the nonlinear model, RMSE value is used like that used in the linear regression except the difference that the method for minimizing the SSE is based on iterative nonlinear least square regression [25, 26], at which point the normal equations for the nonlinear models comprise of partial derivatives of the nonlinear function for each of the regression parameters. Herein, a nonlinear model of the following form was used:

$$T = a(B^{n_1})\,(P^{n_2})\,(D^{n_3}) \tag{9}$$

From Table 5, it can be seen that the RMSE has been significantly reduced to 434.82 Cumecs as compared with 452.277 Cumecs of the best fit-linear model (Eq. (7)), whereas the N-SE has been increased to 97.15% from 96.9%, indicating that the nonlinear model gave a better performance as compared with the linear models. Further, D is the least influential as indicated by the CI of $n_3$, whose estimate is outside the range of 95% CI, suggesting that D should be terminated from the model; consequently, the following new model was constructed:

$$T = a(B^{n_1})\,(P^{n_2}) \tag{10}$$

From Table 6, it can be observed that although all the parameters have their estimates in the range of their corresponding 95% CI, the model was further modified since the values of RMSE and N-SE were unaltered. The model was modified by adding an intercept and is given in Eq. (11).

$$T = a_1(B^{n_1})\,(P^{n_2}) + a_2 \tag{11}$$

It was found (see Table 7) that by adding an intercept ($a_2$), a better model came up in terms of its least RMSE and enhanced N-SE of 428.379 Cumecs and 97.23%, respectively. This argument is also supported by the fact that a real-life process can include predictors that are not only multiplied but also accompanied by additive terms (an intercept). Hence, the additive term will also compensate for the inflows from Daggar. The higher value of N-SE indicated that the simulated vs. observed values are now close to the 1:1 line. This can also be seen from Fig. 5 that the deviation between the trend line and the 1:1 line has been significantly reduced as compared to that noted in Fig. 4.

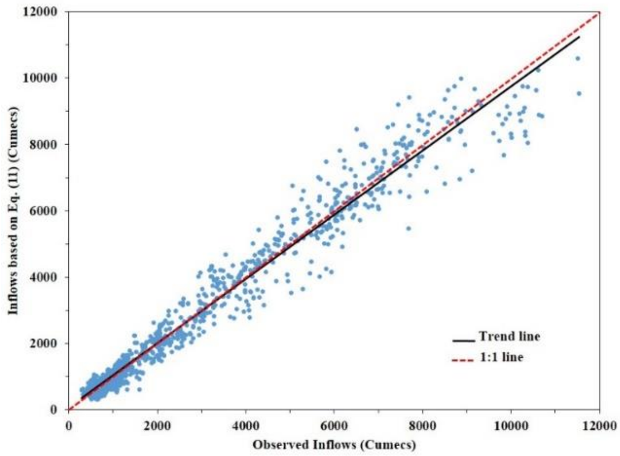


Figure 5: Scatter plot for comparing inflows from Eq. (11) and the observed inflows

To have a visual intuition of the modeled and observed data, the time series plot (see Fig. 6) was developed showing the inflows based on the best-fit nonlinear model given by Eq. (11) and the observed inflows.

Table 5: Regression results of Eq. (9)

| Parameter | Estimate | 95% CI |
|---|---|---|
| A | 1.13704 | (1.00797, 1.28002) |
| $n_1$ | 0.92200 | (0.90780, 0.93643) |
| $n_2$ | 0.12671 | (0.11150, 0.14186) |
| $n_3$ | -0.00147 | (-0.01152, 0.00847) |
| S = 434.829 | E = 0.971 | |

Table 6: Regression results of Eq. (10)

| Parameter | Estimate | 95% CI |
|---|---|---|
| A | 1.13841 | (1.00974, 1.28090) |
| $n_1$ | 0.92194 | (0.90777, 0.93635) |
| $n_2$ | 0.12572 | (0.11207, 0.13930) |
| S = 434.532 | E = 0.971 | |

Table 7: Regression results of Eq. (11)

| Parameter | Estimate | 95% CI |
|---|---|---|
| $a_1$ | 0.533 | (0.408, 0.693) |
| $n_1$ | 1.003 | (0.974, 1.033) |
| $n_2$ | 0.126 | (0.112, 0.140) |
| $a_2$ | 202.859 | (143.296, 260.423) |
| S = 428.379 | E = 0.972 | |

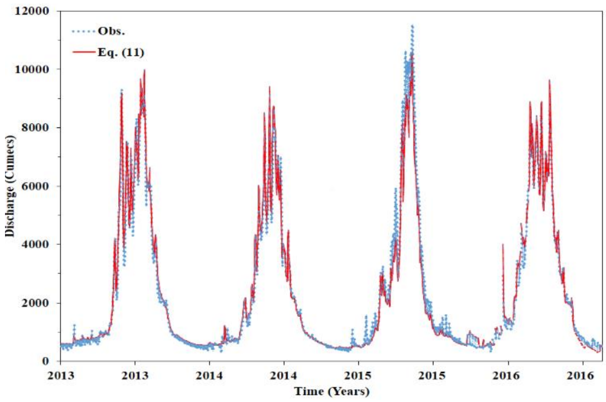


Figure 6: Time series plot for comparing the fitted inflow values of Eq. (11) with observed inflow values

Fig. 6 illustrates that the modeled inflows closely matched the observed data over the four years from 2013 to 2016. However, it slightly underestimates the observed inflows at some peaks for the year 2015. This could be because of the measurement errors usually occurring at the gauging stations of Besham [10, 13]. Because of the discussion based on the regression analysis of all the testing models, Eq. (11) is chosen as the best-fit model for estimating the inflows to the Tarbela reservoir.

## 5 Conclusions

Discharge measured at Besham Qila is the most influential predictor whereas that measured at Daggar is the least influential predictor for the assessment of inflows to Tarbela reservoir. The proposed model, Eq. (11), has proved to be the best-fit nonlinear model since it yielded the least Root Mean Square Error (RMSE = 428.37 Cumecs) and highest Nash-Sutcliffe Efficiency (N-SE = 97.23%), whereas the existing model given in Eq. (2) used by TDP resulted in higher Root Mean Square Error (RMSE = 657 Cumecs) and lower Nash-Sutcliffe Efficiency (N-SE = 93.48%). The proposed model is given in Eq. (11) became helpful in resolving the conflict between Indus River System Authority (IRSA) and Tarbela Dam Project (TDP) Engineers for better management of Indus water. (Note that the existing model being used by TDP Engineers (Eq. (2)) was overestimating the inflows to Tarbela reservoir by 10%). It is recommended that the hydrologic model should be modified after every three to four years based on the latest observed inflows data at Tarbela dam.

## Acknowledgment

The authors are grateful to the Dam Safety Organization (DSO) WAPDA, Surface Water Hydrology Project (SWHP) WAPDA, and Tarbela Dam Project (TDP) WAPDA for providing necessary data for the execution of the research work.

## Author Contributions

Noor Yaseen analyzed the data, developed new models, and carried out model validation. He wrote the manuscript and took primary responsibility for communicating with the journal. Habib ur Rehman supported the documentation process and supervised the research work. Salik Haroon Abbasi arranged the data and supported the documentation process.

## Competing Interests

The authors declare that there is no conflict of interest.

## References

1. Todini, E. Hydrological Catchment Modelling: Past, Present and Future. Hydrology and Earth System Sciences, 2007. 11, 468-482. DOI: http://dx.doi.org/10.5194/hess-11-468-2007.
2. Adamowski, J., & Chan, H. F. A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology, 2011. 407(1-4), 28-40.
3. Maheswaran, R., Adamowski, L., Khosa, R. Multiscale streamflow forecasting using a new Bayesian model average based ensemble multi-wavelet Volterra nonlinear method. Journal of Hydrollogy, 2013. 507, 186-200.
4. Nourani, V., Baghanam, A., Adamowski, J., Kisi, O. Applications of hybrid wavelet-artificial intelligence models in hydrology: A review. Journal of Hydrology, 2014. 514, 358-377.
5. Devi, G.K., Ganasri, B.P. and Dwarakish, G.S. A Review on Hydrological Models. Aquatic Procedia, 2015. 4, 1001-1007. DOI: http://dx.doi.org/10.1016/j.aqpro.2015.02.126.
6. Pizarro-Tapia, R., Valdés-Pineda, R., Olivares, C. and González, P.A. Development of Upstream Data-Input Models to Estimate Downstream Peak Flow in Two Mediterranean River Basins of Chile. Open Journal of Modern Hydrology, 2014. 4, 132. DOI: http://dx.doi.org/10.4236/ojmh.2014.44013.
7. Roca., M. Tarbela Dam in Pakistan. Case study of reservoir sedimentation. Published in the proceedings of River Flow, 2012. 5-7.
8. WAPDA. Tarbela dam 4th periodic inspection report: Sedimentation management study. Pakistan Water and Power Development Authority, Lahore. 1998.
9. Tarbela Dam Project, WAPDA. Report on fifth periodic inspection, 2007. 1, 160.
10. Integrated Consulting Services. Report on sixth periodic inspection, Dam Safety Organization, WAPDA, Lahore, 2018. P.315.
11. Ullah, S. F. Time Series Forecasting with Artificial Neural Network of Water Outflow in Tarbela Dam, Pakistan (Doctoral dissertation, University of Balochistan, Pakistan) 2020.
12. Tu, J. V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of clinical epidemiology. 1996, 49(11), 1225-1231.
13. Sanner, H., Harlin, J., & Persson, M. Application of the HBV model to the Upper Indus River for inflow forecasting to the Tarbela dam. SMHI, 1994. 48.
14. WAPDA. Tarbela Dam Project, Residency Survey and Hydrology Organization; Water and Power Development Authority: Lahore, Pakistan, 2017.
15. Amin, A., Iqbal, J., Asghar, A., & Ribbe, L. Analysis of current and future water demands in the Upper Indus Basin under IPCC climate and socio-economic scenarios using a hydro-economic WEAP model. Water. 2018, 10(5), 537.
16. Rehman, S. A., Bui, M. D., Hasson, S., Rutschmann. P. An Innovative Approach to Minimizing Uncertainty in Sediment Load Boundary Conditions for Modelling Sedimentation in Reservoirs. Water, 2018. 10, 1411. DOI: http://dx.doi.org/10.3390/w10101411.
17. Ahmad, Z., Hafeez, M., Ahmad, I. Hydrology of mountainous areas in the upper Indus Basin, Northern Pakistan with the perspective of climate change. Environmental Monitering and Assessment, 2012. 184(9), 5255-5274. DOI: http://dx.doi.org/10.1007/s10661-011-2337-7.
18. TAMS Consultants and HR Wallingford. Tarbela Dam Sediment Management Study; Main Report; Water and Power Development Authority: Lahore, Pakistan, 1998.
19. Surface Water Hydrology Project. Sediment Appraisal of Pakistan Rivers 1960-1998. Surface Water Hydrology Project, Hydrology and Water Management Organization WAPDA, Lahore, SWHP Publication, 2001. 57, P.222.
20. Spiegel, M. R., and Stephens, L. J. Theory and Problems of Statistics, Fourth Edition, the McGraw-Hill Company, 2008. P.601.

21. Chatterjee, S., Hadi, A.S. Regression analysis by example, Fourth Edition, John Wiley & Sons, Inc., Hoboken, New Jersey, 2006. P.366.
22. Legates, D. R., & McCabe Jr, G. J. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research, 1999. 35(1), 233-241.
23. Nash, J.E. Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology. 1970, 10, 282–290.
24. Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. Probability and statistics for engineers and scientists. New York: Macmillan, 1993. 5, P.812.
25. Ott, R. L., & Longnecker, M. T. An introduction to statistical methods and data analysis. Sixth Edition, Nelson Education, 2015. P.1296.
26. Brown, A. M. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet. Computer Methods and Programs in Biomedicine, 2001. 65, 191-200.