



OLS regression based causal modelling for flood management: A study on river Mayurakshi at Tilpara Barrage section, Suri, Birbhum

R. Biswas*, S. Pal, S. Bhattacharyya and K. Mandal

National Institute of Technology Durgapur, Durgapur-713 209, West Bengal, India

E-mail: raktim.biswas6215@gmail.com

Manuscript received online 28 March 2020, revised and accepted 20 September 2020

Flooding is one of the reasons of endangering human and civil life in the Mayurakshi river basin at Tilpara Barrage area under Suri Block I, in the district of Birbhum, West Bengal, India. Number of efforts have been made by different groups of researchers including researchers in the field of engineering to develop technology as to analytically manage and mitigate the same. Existing research work have identified variables namely 'rainfall', 'inflow', 'canal discharge' and 'upstream water level' are cause behind the severity of devastation created by flood. However, no such effort has been employed earlier to identify statistically significant relationship among the four variables. Present research has used 20 years' time series data vis-à-vis these four variables and employed OLS regression techniques to develop a model for flood management. This model also explains the relative importance of each variable on other and overall in flood management. Finally, this model is a step forward controlling flood which accounts huge loss of life, properties and life stock.

Keywords: River Mayurakshi, Tilpara Barrage control structure, OLS regression modelling, flood management.

Introduction

Flood is such a natural phenomenon that is uncontrollable and, to some extent, unpredictable. Man must acquaint himself with the characteristics of floods if he is to control them. Floods vary as the weather differs from month to month and year to year. The first step in becoming better acquainted with flood is to measure them. When this has been done and record is available for a long period, analytical studies will give good explanation of the event (Chow *et al.* 1964). Floods in the basin happened due to unusual raining in monsoon time (Ghosh *et al.* 2015). Former studies of flood discharge in the basin indicate that probability of flooding is affected due to occurrence of flood after construction of Tilpara Barrage in the river basin. According to Jha and Bairagya (2012), the Massanjore dam influences flow to the Tilpara Barrage and it affects high flows of the Mayurakshi river basin. Bhattacharya and Let (2012, 2013) indicate connection of inundation and discharge from controlled section. Record of high flow rate and its occurrence in the basin affect the flow record of the barrage. Jha and Bairagya (2012) studied on flood phenomenon of this river basin. In earlier works, ANN

(Artificial Neural Network) modelling has been done but it is found that for hydrological analysis of river basin (in case of ungauged basin) for hydrological extreme events like flooding the classical statistical process regression performs much more efficiently. Lee *et al.* (2012) applied individual bivariate probability models to map flood susceptible areas in Busan, South Korea considering the flood occurrence variables (dependent and independent). Pradhan *et al.* (2010) utilized multivariate logistic regression to get relationship between two kinds of variables. Tehrany *et al.* (2013) combined these methods for flood susceptibility analysis in tropical regions and enhance the combined method over individual method. Regression analysis is a tool to devise the statistical flow control of the river section. Barrage section is composed of the conventional control structures installed therein at the river gauging station. The barrage structure composed of mainly: u/s river section, d/s river section, CHR (Canal Head Regulator), sluice gate and valve etc. After the average supply to the Mayurakshi canals as in MDMC and MBMC (Mayurakshi Dwaraka Main Canal and Mayurakshi Bakreswar Main Canal) for the agricultural purpose, the remaining river

water is discharged through the barrage gate in the downstream of river course. In the upstream from Massanjore Dam at Jharkhand, released river-water flows in the main course and its tributaries and come to Tilpara Barrage at Suri and then the water after the barrage releases, flows by the respective river course and heads towards the Abhedananda bridge section near Sainthia in Birbhum district. The river course of Mayurakshi is filled with sand bar, braided bar channel, drainage pocket etc. with varying flora and fauna in the river course and in high flood situation.

Description of the study area

The river Mayurakshi is a 5th order tributary of the river Bhagirathi. Its catchment area lies within the transitional zone between two mega physiographic provinces namely Chotonagpur plateau and the Bengal basin. The Mayurakshi originates from Dumka district in Jharkhand and then flows through Birbhum district of West Bengal, bisecting Suri subdivision of Suri Block I. Mayurakshi has a length of 288 km and a basin area of 5325 sq. km, spread over coordinate values as 23°15'N-24°34'15"N Latitude and 86°58'E-88°20'30"E Longitude (Fig. 1). It starts from the foot of

Trickutpahar hills of Jharkhand. Branches and channels as Manikornika, Gambira, Kana Mayurakshi, Mor, Beli or Tengramari etc. form channel-network flows to Hizol-beel, from there onwards as river Babla, and finally joins Bhagirathi. The river basin specially the lower part is a well-known name in flood scenario of West Bengal.

Research methodology in nutshell

Existing research gap and objective of present research:

Since, it is important to estimate the devastating intensity of flood, from the existing literature and from domain knowledge of the concerned subject, we have identified variables 'upstream water level in feet', 'total canal discharge', 'inflow acre-feet per day' and 'rainfall' and their multivariate interplay is the cause behind occurrence of flood or quasy flood situations. In spite of this identified reason no research has been undertaken to develop a relational model among these stated variables. Present research has been undertaken for mitigate this research gap. Moreover, this research has not only put the focus to identify the statistically significant relationship among the stated variables but also has emphasized further to understand relative importance of each

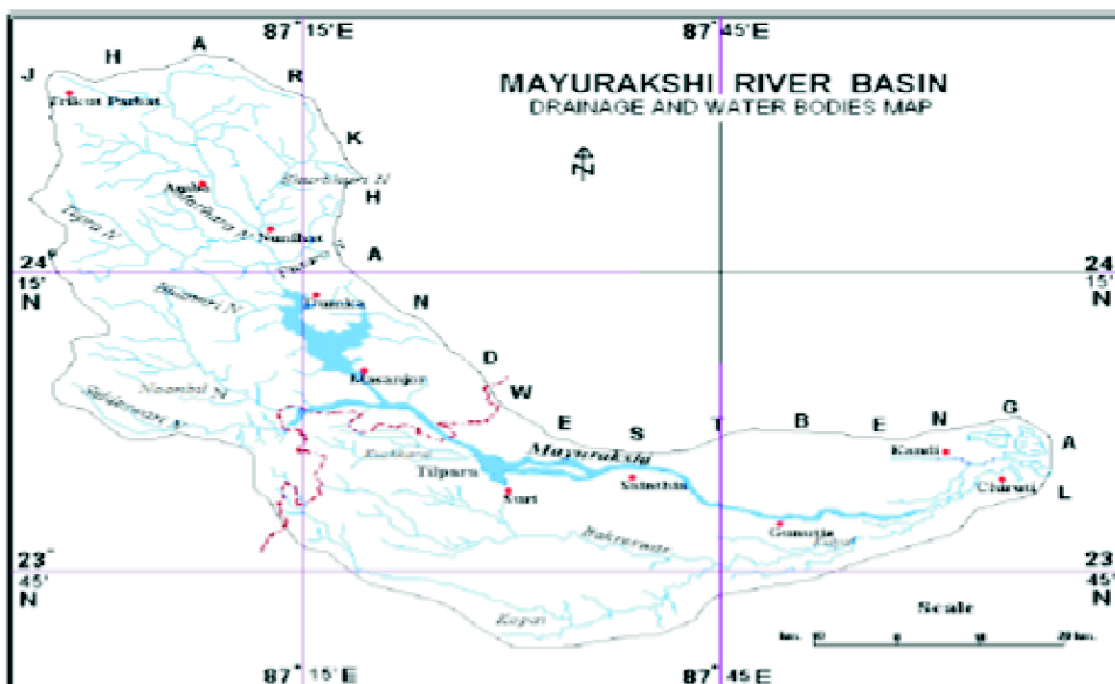


Fig. 1. Mayurakshi River Basin (MRB).

Table 1. Data of Tilpara Barrage: Source: <http://www.wbiwd.gov.in>

Location	Suri, Birbhum, river Mayurakshi
Basin area	3208 sq. km
Abutment width	308.84 m
Bays	7 (18.29 m), Sluices 8 (18.29 m)
Water way	274.39 m
Pond level	64.33 m
Upstream water level	64.63 m
Discharge	8,495 m ³ /s
Canal discharge	99.11 m ³ /s
Main canal	Left 16.62 km , Right 22.53 km
Irrigation area	Kharif 226,629 Ha; Rabi 20250 Ha
Irrigation achieved	Kharif 220730 Ha; Rabi 8150 Ha; Boro 25400 Ha

variable on other. Therefore, present research has two prong objectives, one, is to identifying statistically significant relationship and other is to unfold relative importance of each variable in the flood management.

Data sources and nature of data:

The daily data of river at Tilpara from 1996 (August)-2016 (October) have been obtained from the Irrigation and Waterways Department, Suri-Birbhum, West Bengal in the form: 'Date, Pond level, Down-stream level, Inflow, Rainfall, Supply to canals: Mayurakshi Bakreswar Main Canal (MBMC) and Mayurakshi Dwaraka Main Canal (MDMC), Total supply, Average supply to canals: MBMC and MDMC and River discharge, Average discharge and Maximum discharge' etc. The collected original data has been converted to excel format. Then the whole data is reduced to the duration of 'Monsoon Period' (1st June to 31 October) for each year and has been accumulated. From this total set of data, only 'pond level-inflow-rainfall-canal discharge (total)' is taken and the final data-set were ready for analysis. The analysis is done in SPSS and MS Excel (2007) Office Package. And for each data-set, the graphical plot and tables is drawn against each data-range for each data-points. In relation to the research query developed appropriate inferential statistics like OLS regression and necessary descriptive statistics like 'Skewness', 'Kurtosis', 'Maximum', 'Minimum', 'Standard Deviation' have been employed.

Methodology:

In the previous section, theoretical discussion has helped

us to construct a proposed model of flood management. This model conceptually developed some relations that need to be empirically verified. Relationships conceptually established can be presented in the following as proposed hypotheses:

H_{A1}: 'Rainfall' is related to 'Inflow'

H_{A2}: 'Rainfall' is related to 'Upstream water level'

H_{A3}: 'Inflow' is related to 'Upstream water level'

H_{A4}: 'Inflow' is related to the 'Total canal discharge'

H_{A5}: 'Upstream water level' is related to 'Total canal discharge'

These hypotheses constitute a proposed model (see in the Diagram 1) for flood management and that is required to be confirmed empirically. This whole is our first research query.

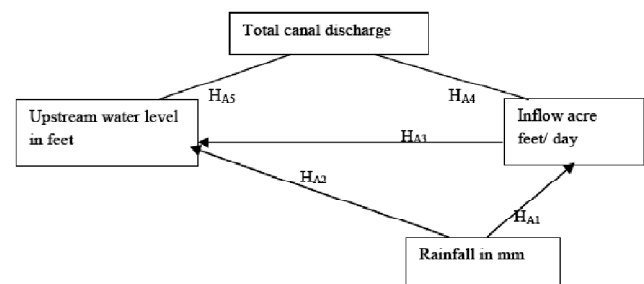


Diagram 1. Proposed model.

In this present research, we are not only concerned of the establishing stated relationships but also our other interest is to study and empirically measure the relative importance of all the anchors in managing flood and it forms our second research query.

And it is not out of place to mention that this present research is employing statistical significance of measurement based on available data. Further, in the present work we also have third set of research query and that is related to an analysis, whether or not measured relative importance is stable in between the periods. This would help us to understand the stability of the model across the years and may have helped to conceptualise the cause if it is not so. In relation to placed research queries, we have decided to develop a two layers linear model based on Ordinary Least Square Regression (OLSR). In layer one, we consider 'Rainfall' as

independent variable and Inflow and Upstream water as dependent variables for two separate OLSR. Simultaneously, taking Inflow as independent variable, upstream water level as dependent variable another OLS Regression model has been executed. In layer two 'upstream water level' and 'Inflow' have been considered as two independent variables and 'Total canal discharge' as dependent variables, a OLS Regression has been carried out. For the purpose of empirical testing with the help of a OLSR we have deployed secondary data collected from Mayurakshi river at Tilpara Barrage for said four variables for last available 20 years (1996-2016). Using the result of OLSR we can able to test the hypothesis and able to develop the final model and that would be fulfilment of our research query 1. For the purpose of getting answer of the research query 2, we have decided to use standardised regression coefficient. First of all, regression coefficient is a measure of change of one unit of independent variable to the amount in change in the dependent variable when, if other independent variable present, remain constant or unchanged. Thus it is a measure of impact of one variable over other. Further statistical testing has been carried out to understand the amount of impact is significant (away from term) or not. But in OLSR there are two coefficients, one is standardised and another is un-standardised, so question may arise which is right for our model? Since there four deployed variables are of different units therefore to overcome this as a problem we imply standardized regression coefficient because it is based on unit free standardised normal variate. To answer the research query two, we have measured the percentage of relative importance of each independent variable in case where more than one variable is working as independent variables. We used following formula for the purpose. When independent variables are V1 and V2 and relative importance of V1 and V2 are written as RI of V1 and in case of V2 is written as RI of V2.

$$\text{RI of V1} = \frac{\text{Coefficient of V1}}{\text{Coefficient of V1} + \text{Coefficient of V2}}$$

$$\% \text{RI of V1} = \frac{\text{Coefficient of V1}}{\text{Coefficient of V1} + \text{Coefficient of V2}} \times 100$$

Similarly,

$$\text{RI of V2} = \frac{\text{Coefficient of V2}}{\text{Coefficient of V1} + \text{Coefficient of V2}}$$

$$\% \text{RI of V2} = \frac{\text{Coefficient of V2}}{\text{Coefficient of V1} + \text{Coefficient of V2}} \times 100$$

Here coefficient of V1 and V2 are the Value of standardised regression coefficient of independent variable 1 and 2. And percentage of RI of V1 or V2 is just multiplication RI with 100. Addressing the third research query we have segregated the whole data set (20 years) into four different sub-sections (case1) and also with two sub-sections (case 2) with an objective to re-executed OLSR for each section and re-examined the value of standardised regression coefficient and statistical significance of its impact for each sub-section. Further, we have compared the significant values of SRC across four subsections and two sub-sections respectively for case 1 and case 2.

Results and discussion

The standard tables generated are given below. For the data series of upstream water level, inflow, rainfall, total canal discharge and maximum river discharge; the statistical parameters like maxima, minima, standard deviation. Skewness, kurtosis was calculated. At end of table, Pearson coefficient is calculated treating rainfall as independent variable and all other variables as dependent variables. The coefficient bears a positive co-relation with inflow of the river (u/s) section and river discharge at (d/s) section and co-relate negatively with (u/s) water level and total canal discharge (u/s). All the other values in calculating coefficient, gives fractional values (-ve or +ve) when rainfall value is treated independent variables of the functions (Table 2).

Essential diagrams or plots generated are given below. In the plot, the ordinate is from 0 to 10,000 to 100,000; its interval having 10,000 of data points (of flow volume) and in abscissa, the plot yields data point (each against) corresponding to data of Upstream Water Level (feet), Inflow (Acre-feet), Rainfall (mm) and Total Canal Discharge (Cubic feet/s) marked separately with legend in the box graphical plot. The graph shows a usual pattern of data against the four vari-

Table 2. Statistical estimate of Mayurakshi river data series

Excel river data set (1992-2016)	Max.	Min.	Std. dev.	Skewness	Kurtosis
Upstream water level	206.1	196.2	1.06	-1.06	2.7
inflow	91525	0.01	6888.77	4.54	35.41
Rainfall	140.6	0	14.84	3.40	14.80
Total canal discharge	55500	49	2783.05	9.6	149.75
Max. river discharge	65599	0	5383	5388.09	34.98

ables (shown in figure). The combined graph, shows a rising in pattern corresponding to abscissa point (1475,1609) and takes ordinate value from 60,000 to 90,000. In case of up-stream water discharge graph, it gives for the same value of abscissa points and shows a decrease in value at point 195 (ft) in the plot or diagram.

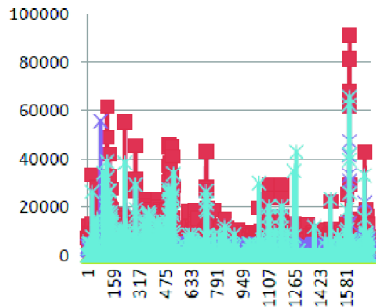


Fig. 2. Combined graph.

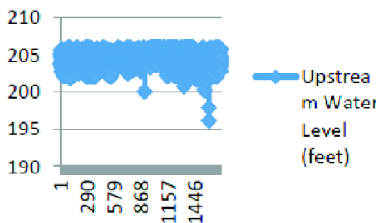


Fig. 3. Upstream water level (feet).

Statistical analysis of the hypotheses of the proposed model of flood management:

We have received the results of five proposed hypotheses after executing ordinary least square regression. Two hypotheses got rejected (see Table 3 for summary of results) and that result in reconstruction of the proposed model. Our theoretical proposal that 'Rainfall' is having its impact on upstream water and 'Inflow' is creating its influence over the

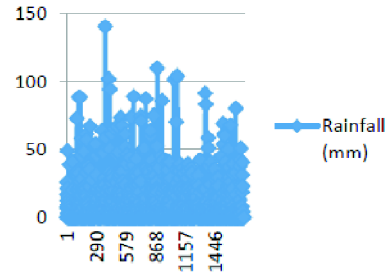


Fig. 4. Rainfall (mm).

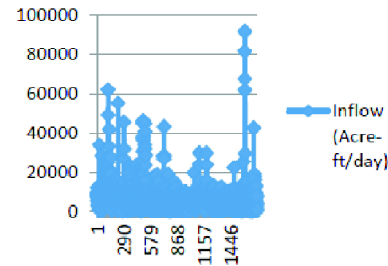


Fig. 5. Inflow (acre-ft/day).

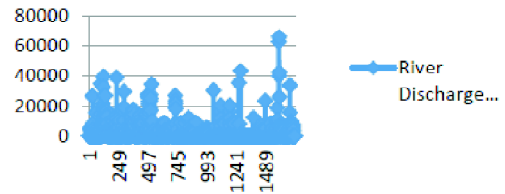


Fig. 6. River discharge downstream (cubic ft/s)

upstream water, were not found significant statistically (empirically). Thus, final model is linked in such that 'Rainfall' is affecting 'inflow' and inflow is affecting total canal discharge alongside upstream water level (see Diagram 2). Since standardised Beta between rainfall (as independent variable) and inflow (as dependent variable) is 0.128 with probability of significance is 0.00%, we can consider 1 unit of increment of 'rainfall' may increase 'inflow' 0.128 unit. Further, simultaneously 'upstream water level' and 'inflow' are having their impacts on 'total canal discharge'. Standardised Beta are 0.190 and 0.511 respectively for the same. And these have denoted one unit change of each when other remains constant may have increased total canal discharge 0.190 and 0.511 unit respectively. Thus relative importance of 'upstream water level' and 'inflow' are respectively 27% and 73% (approximately) for influencing total canal discharge; which is

Table 3. Summary of hypotheses

Sr. No.	Hypotheses proposed	Remarks
1.	H _{A1} : 'Rainfall' is related to 'Inflow'	Accepted with 99% CI for the period 1997-2016, 2012-2016, 1997-2006 and 2007-2016. But accepted with 95% CI for the period of 1997-2001, 2007-2011
2.	H _{A2} : 'Rainfall' is related to 'Upstream water level'	Rejected
3.	H _{A3} : 'Inflow' is related to 'Upstream water level'	Rejected
4.	H _{A4} : 'Inflow' is related to the 'Total canal discharge'	Accepted with 99% CI for period 1997-2016 and also for 1997-2001, 2002-2006 and 2007-2011 and 1997-2006
5.	H _{A5} : 'Upstream water level' is related to 'Total canal discharge'	Accepted with 99% CI for period 1996-2016 and also for 1997-2001, 2002-2006 and 2007-2011 and 1997-2006

Table 3A. OLSR output for layer one (period 1996-2016)

Goodness of model fit	R ²	0.288	Significant probability value of F ratio	0.000
Dependent variable	Total canal discharge		Standardised Beta	Probability value of significance
Independent variable	Upstream water level in feet	0.190		0.000
	Inflow acre feet/day	0.511		0.000

Table 3B. OLSR output for layer two (period 1996-2016)

Goodness of model fit	R ²	0.016	Significant probability value of F ratio	0.000
Dependent variable	Inflow acre feet/day		Standardised Beta	Probability value of significance
Independent variable	Rainfall in mm	0.128		0.000

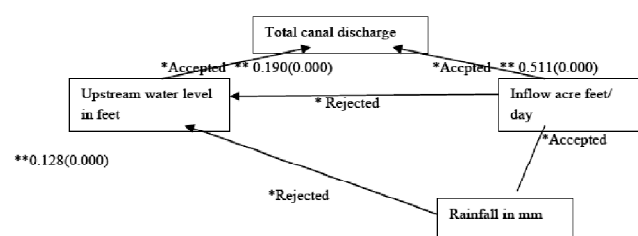


Diagram 2. Accepted model or final model.

*Rejection or acceptance mean statistically rejection or acceptance of proposed hypothesis.

**Standardised Beta value (probability of significance)

the key anchor for flood in the neighbourhood area. Further, to test the empirical stability of the proposed model we have analysed the results after dividing the data-set in two and four equal sub-sections. Results have shown a clear variation of uniformity (stability) of the model. If we follow the Table 4 we can observe relative importance of upstream water is reduced between period (1997-2001) and period (2000-2006) whereas it again increased in the period (2007-2011) but reduced so less to statistically insignificant in the period (2012-2016). Just the reverse happened for 'Inflow'. Similar observation found when we have divided the data-set in two sub-sections (1997 to 2006) and (2007-2016) (see Table 5). In

Table 4. OLSR output for layer one and two across four time sections

Period	Dependent variable	1997-2001		2002-2006		2007-2011		2012-2016	
		B value	SV (Significant value)	B value	SV	B value	SV	B value	SV
Independent variable	Upstream water level in feet	0.653	0.000	0.311	0.000	0.463	0.000	(-) 0.016	0.320
	Inflow acre feet/day	0.208	0.000	0.478	0.000	0.300	0.000	0.961	0.000
Goodness of fit	R ²	0.423	x	0.392	x	0.456	x	0.93	x
	Probability value of F	0.000	x	0.000	x	0.000	x	0.000	x

Table-4 (contd.)

Dependent variable	Inflow acre feet/day	Standardised Beta	Significant probability value	B value	SV	B value	SV	B value	SV
Independent variable	Rainfall	0.131	0.018	0.079	0.127	0.110	0.028	0.207	0.000
Goodness of fit	R^2	0.017	x		0.006	0.012		0.043	x
	Significant probability value	0.018	x		0.127	0.028		0.000	x

Table 5. OLSR output for layer one and two between two time sections

Period		1996-2006		2007-2016	
		B value	SV	B value	SV
Dependent variable	Total canal discharge				
Independent variable	Upstream water level	0.524	0.000	(-) 0.034	0.040
	Inflow	0.295	0.000	0.900	0.000
Goodness of fit	R^2	0.338	x	0.81	x
	Significant value of F ratio	0.000	x	0.000	x
DV	Inflow	B value	SV	B value	SV
IV	Rainfall	0.121	0.001	0.157	0.000
Goodness of fit	R^2	0.015	x	0.025	x
	Significant value of F	0.001	x	0.000	x

case of layer two of the model where 'Rainfall' is considered as independent variable and Inflow has considered as dependent, between them variable, it has been observed that almost all periods except the period 2002-2006 and found significant relationship and significant standardised Beta. And same is also happened for two sub-sections i.e. period 1997-2006 and period 2007-2016 (see Table 5).

Conclusions

After successful completion of statistical analysis, it can be concluded that considering a flow coefficient over the barrage section (i.e. Pearson correlation coefficient), considering rainfall, Inflow and upstream water level as variable or independent over the data analysis and downstream water level as a dependent (as from field data or same as in table-plot), with the changes (in terms of increment or decre-

ment) in the values of rainfall amount, the amount of inflow and pond level over the river cross-section and subsequently total canal discharge supplied through also change (increases or decreases) affects the flow control or regulation of Tilpara Barrage over high or peak flow or high flood discharge situation in monsoons in rainy season in August-September-October. The main combined graph is shown in Fig. 2 including the peak values or rise in the graph clearly shows the influence of rainfall in flood probability of Mayurakshi river basin with the help of previous irrigation data-sheet which regulates flow modulations within the gate or sluice valves of the barrage section of the river water course respectively.

Acknowledgements

Research work is supported by Irrigation and Waterways Department, Suri-Birbhum, West Bengal, India.

Appendix

Table 6. Canal data model 1997-2001

R^2	0.423	F value	119.097	Sig. probability value of F	0.000
Dependent variable	Total canal discharge in cubic feet per second	B value	Beta value	t statistics	Significant probability value
Constant	-	247462.736	-15.142	0.000	
Independent variable 1	Upstream water level in feet	1220.585	0.653	15.266	0.000
Independent variable 2	Inflow acre-feet per day	0.045	0.208	4.861	0.000

Table 7. Canal data model 2002-2006

R^2	0.392	F value	118.908	Sig. probability value of F	0.000
Dependent variable	Total canal discharge in cubic feet per second	B value	Beta value	t statistics	Significant probability value
Constant	–	122240.355	–	–7.404	0.000
Independent variable 1	Upstream water level in feet	603.579	0.311	4.475	0.000
Independent variable 2	Inflow acre-feet per day	0.171	0.478	11.485	0.000

Table 8. Canal data model 2007-2011

R^2	0.456	F value	164.674	Sig. probability value of F	0.000
Dependent variable	Total canal discharge in cubic feet per second	B value	Beta value	t statistics	Significant probability value
Constant	–	–123958.065	–	–10.292	0.000
Independent variable 1	Upstream water level in feet	615.497	0.463	10.420	0.000
Independent variable 2	Inflow acre-feet per day	0.109	0.300	6.753	0.000

Table 9. Canal data model 2012-2016

R^2	0.931	F value	1941.920	Sig. probability value of F	0.000
Dependent variable	Total canal discharge in cubic feet per second	B value	Beta value	t statistics	Significant probability value
Constant	–	11294.782	0.987	0.324	
Independent variable 1	Upstream water level in feet	–55.792	–0.016	–0.997	0.320
Independent variable 2	Inflow acre-feet per day	0.480	0.91	60.705	0.00

References

1. V. Bolshakov, *Information Technology and Management Science*, 2013/16.
2. V. T. Chow, D. R. Maidment and L. W. Mays, "Handbook of Applied Hydrology", International Edition, Mc Graw Hill Company, New York, 1964.
3. M. Cupal, O. Deev and D. Linnertova, *Procedia Economics and Finance*, 2015, **23**, 499.
4. N. Dogulu, L. Lopez, D. P. Solomatina, A. H. Weerts and D. L. Shrestha, *Hydrology and Earth System Sciences*, 2015, **19**, 3181.
5. K. G. Ghosh and S. Mukhopadhyay, *Eastern India Asian Research Publishing Network (ARPN)*, 2015, **4(2)**.
6. K. G. Ghosh, S. Mukhopadhyay and S. Pal, *International Research Journal of Earth Sciences*, 2015, **3(7)**, 11.
7. K. G. Ghosh and S. Pal, *European Water*, 2015, **50**, 3.
8. V. Griffis V and J. R. Stedinger, *Journal of Hydrology*, 2007, **344**, 82.
9. V. C. Jha and H. Bairagya, Flood Plain Planning Based on Statistical Analysis of Tilpara Barrage Discharge, A case study on Mayurakshi River Basin; <http://www.seer.ufu.br/index>.
10. S. Mishra, C. Saravan, V. K. Dwivedi and J. P. Shukla, *Journal Geographical Society of India*, 2018, **92**, 305.
11. S. Mukhopadhyay and S. Let, *Trans. Inst. Indian Geographers*, 2018.
12. S. Pal and S. Let, *International Journal of Advanced Research in Management and Social Sciences*, 2018.
13. S. Pal, S. Let and P. Das, *International Journal of Geology, Earth and Environmental Sciences*, 2012, **2(3)**.
14. B. Pradhan, *Journal of Spatial Hydrology*, 2009, **9(2)**.
15. M. Rezaeanzadeh, H. Tabari, A. Yazdi, S. Isik and L. Kalin, "Flood flow forecasting using ANN, ANFIS and regression models", *Neural Comput & Applic*, Springer-Verlag, 2013.
16. M. S. Tehrany, M. J. Lee, B. Pradhan, M. N. Jebur and S. Lee, *Environ. Earth Sci.*, 2014, **72**, 4001.