DEVELOPMENT OF STOCHASTIC MODELS FOR THREE SUB-CATCHMENTS OF TEHRI DAM

A DISSERTATION REPORT

Submitted in partial fulfilment of the requirements for the award of the degree

of

MASTER OF TECHNOLOGY

in

HYDROLOGY

by

ALLY DIWANI

DEPARTMENT OF HYDROLOGY

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

ROORKEE – 247 667 (INDIA)

May 2019

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in this report entitled, '**DEVELOPMENT OF STOCHASTIC MODELS FOR THREE SUB-CATCHMENTS OF TEHRI DAM'**, in partial fulfilment of the requirements for award of the degree of **Master of Technology in Hydrology** submitted to the Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out under the guidance of **Dr N.K. Goel, Professor, Department of Hydrology, IIT Roorkee** during the period from July 2018 to May 2019.

The matter embodied in this dissertation has not been submitted by me for the award of any other degree.

Date: …...May, 2019

Ally Diwani,

Enrolment number 17537001

CERTIFICATE

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dr N.K. Goel,

Professor,

Department of Hydrology,

Indian Institute of Technology Roorkee,

Roorkee – 247 667 (Uttarakhand), India.

ACKNOWLEDGEMENTS

First of all, I would like to thank God Almighty for his blessings to enable me to complete my studies and empower me to complete my research work. Without his blessings, this achievement would not have been possible.

With a very special sense of gratefulness, I would like to express my thanks to my supervisor **Dr N.K. Goel, Professor Department of Hydrology** for his dedicated assistance, willingness to give his time, comments, and constructive suggestions throughout the work.

I would like to express my heartfelt thanks to Dr M. K. Jain, Head, Department of Hydrology, Dr H. Joshi, Dr M. Perumal, Dr D. S. Arya, Dr B. K. Yadav, Dr Sumit Sen, and Dr Jaya Khanna, Faculty Members, **Department of Hydrology, Indian Institute of Technology, Roorkee** for their excellent guidance, valuable teaching, assistance and encouragement during the entire course of my study at IIT Roorkee.

Further gratefulness goes to **ITEC Programme of Government of India**, which sponsored my study in Master of Technology (Hydrology) programme of IIT Roorkee. Special thanks go to my wife Mrs Violeth John and my sons, for being patient during my stay in India. I also recognize and appreciate the lifelong influences of my parents, siblings and friends whose personal sacrifices showed me the way of achieving my goal.

I express my sincere thanks to **Mr Niraj Kumar Agrawal** Deputy General Manager, THDC India Limited for his support and valuable assistance during the study. The help rendered by Dr Litan Kumar Ray, Rohit Kumar Varshney, and Mr Bhanesh, Project staff is thankfully acknowledged. Also, I would like to thank Ms Bratati (Research Scholar of DOH), all my fellow colleagues and friends for their constant help and encouragement during the course of the study. The help received from the staff members Department of Hydrology and other staff members of IIT Roorkee is thankfully acknowledged.

I would like to thank THDC India Limited for support and allowing me to use the data of the project for use in this study and the cooperation extended during the field visits.

Dated: …... May, 2019 **Ally Diwani**

ABSTRACT

Streamflow forecasting is a crucial step in many of the activities related to planning, management and operation components of water resources systems. Streamflow forecasting is important to the water resources system managers for making proper allocations of water to hydropower generation, irrigation, domestic and other uses on day to day basis. In recent times, due to the effect of changing the climate, the job of water managers has become more important and risky. In a country like in India, where the rainfall occurs mainly during the south-west monsoon months (June to September), the storage and proper utilization of water is a basic need. The development of a proper inflow forecasting system can be very useful for suitable utilization of storage waters. The forecasting of streamflow could be done for short-term as well as for long term basis. In this research, the short term duration of one day has been used for the development of forecasting models.

The main aim of the present study is to develop the stochastic models for three sub-catchments of the Tehri dam. Tehri dam was constructed on the confluence point of Bhagirathi and Bhilangana river, which are one of the sources of great Ganges river of India. The dam is built for multipurpose use. It is the main source of water supply for the Ganga canal and millions of people are dependent on the water supply from the Tehri reservoir. Therefore, the proper utilization of the storage water from the dam is very important for the people living in the command area of the canals which are receiving water from the dam.

To fulfil the objective, at first, the rating curves have been developed for two sub-basins, namely Bhilangana and Balganga of Tehri catchment using method of least squares and ANN technique. Following this, the stochastic models have been developed for three main sub-catchments of Tehri dam. The results of the stochastic models have been compared with the results of HEC-HMS.

For developing the stage-discharge relationships, the data set of 1st June 2016 to 30th November 2018 from two gauging stations, namely Ghansali in Bhilangana river and Sarasgaon in Balganga river have been used. The performance of both the methods have been evaluated using Nash Sutcliffe Efficiency (NSE) and the coefficient of determination (R^2). The results of the analysis show the good performance of both methods. For the method of least squares, the NSE was more than 95% and the coefficient of determination was more than 0.9. However, the efficiency of the ANN method was slightly better than the method of least squares. The RMSE was far less in the case of ANN.

Stochastic models have been developed for three main sub-catchments of Tehri dam, namely Bhagirathi at MBII, Bhilangana at Ghanshali, and Balganaga at Sarasgaon. In the present study four stochastic models namely Autoregressive (AR) model, Autoregressive models with exogenous inputs (ARX), Autoregressive moving average (ARMA) model, and Autoregressive moving average model with exogenous inputs (ARMAX) have been developed and used for daily streamflow forecasting purpose for monsoon and non-monsoon seasons. The rainfall and discharge data from June 2016 to May 15, 2019, for the three sub-basins, namely Bhagirathi at MB II, Bhilangana at Ghansali and Balganga at Sarasgaon were collected from Real-time inflow forecasting system website of Tehri dam. All the developed models were calibrated and validated by dividing the data into two parts. The performance of all the developed stochastic models has been checked using 6 indices namely NSE, RMSE, PBIAS%, R^2, MAE and AIC. The comparison of the results of stochastic with and HEC-HMS model results shows that the performance of selected stochastic models is far better than the HEC-HMS model for the three sites of the Tehri catchment during calibration and validation period. The programs have also been prepared using R-studio version 3.4.3 for the simulation of daily streamflow by stochastic models.

The recommendations made on the basis of the study and scope for future work are listed below:

- The stage-discharge relationship was drawn only using the data from 2016 to 2018, which may not cover the higher flood records and therefore, during the floods, the developed relationship may give lesser value than actual. For this, the relationship could be redrawn in future by using more dataset and a new relationship can be drawn only for flood situation i.e. for higher values of the flood stages.
- In case of the stochastic model, only AR model was developed for non-monsoon season. In future, development of other stochastic models considering the rainfall and temperature are expected to give better results.
- More efforts are required to be put in for increasing the efficiency of the HEC-HMS model with extended data bases. With extended data base, the efficiency of HEC-HMS is expected to improve further.
- The updating of parameters of stochastic models on a daily basis is recommended in future work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF NOTATIONS AND ABBREVIATIONS

The list of the abbreviation used in the thesis study, as follows;

AR	Auto Regressive
ARX	Auto Regressive with exogenous variables
ARMAX	Auto Regressive Moving Average with exogenous variables
ARMA	Auto Regressive Moving Average
ANN	Artificial Neural Network
RMSE	Root Mean Square Error
NSE	Nash Sutcliffe Efficiency
GA	Genetic Algorithm
MT	Model Tree
GEP	Gene Expression Programming
MLR	Multilinear Regression
TDP	Tehri Dam Project
MAE	Mean Absolute Error
PBIAS	Percentage Bias
TS	Time series
RR	Rainfall Runoff
ACF	Autocorrelation Coefficient Function
PACF	Partial Autocorrelation Coefficient Function
HEC-HMS	Hydrologic Engineering Center's Hydrologic Modelling System
HEC- RAS	Hydrologic Engineering Center's River Analysis System
HEC-GeoRAS	Hydrologic Engineering Center's Geospatial River Analysis System
HEC-GeoHMS	Hydrologic Engineering Center's Geospatial Hydrologic Modelling System
DEM	Digital Elevation Model
SCS	Soil Conservation Service
SMA	Soil Moisture Accounting
ANFIS	Adaptive Neural-based Fuzzy Inference System
SWE	Snow Water Equivalent
MOGA	Multi-Objective Genetic Algorithm
MLP	Multilayer Perceptron

SMN	Single Multiplicative Neuron
BP	Back Propagation
GUI	Graphical User Interface
CN	Curve Number
UH	Unit Hydrograph
m	Meter
mm	Millimeter
m^3/sec	Meter Cubic per second
MCM	million cubic meter
MW	Megawatt
Km	Kilometer

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Streamflow forecasting is a crucial step in many of the activities related to planning, management and operation components of water resources systems. Streamflow forecasting is important to the water resources system managers for making proper allocations of water to hydropower generation, irrigation, domestic and other uses on day to day basis. In recent times, due to the effect of changing the climate, the job of water managers has become more important and risky. In a country like in India, where the rainfall occurs mainly during the south-west monsoon months (June to September), the storage and proper utilization of water is a basic need. The development of a proper inflow forecasting system can be very useful for suitable utilization of storage waters. The forecasting of streamflow could be done for short-term as well as for long term basis. In this research, the short term duration of one day has been used for the development of forecasting models.

The main aim of the present study is to develop the stochastic models for three sub-catchments of the Tehri dam. Tehri dam was constructed on the confluence point of Bhagirathi and Bhilangana river, which are one of the sources of great Ganges river of India. The dam is built for multipurpose use. It is the main source of water supply for the Ganga canal and millions of people are dependent on the water supply from the Tehri reservoir. Therefore, the proper utilization of the storage water from the dam is very important for the people living in the command area of the canals which are receiving water from the dam.

1.2 TEHRI DAM

Tehri dam, a multi-purpose dam, is the highest earth and rock fill dam in Asia having a height of 260.5m. The dam has a gross capacity of storage 3540 MCM, with a capacity of power generation 2400 MW. The dam located in the Garhwal hills of Uttarakhand state is about 1.5 km downstream in the confluence of Bhilangana and Bhagirathi rivers. These two rivers are one of the source of the Ganges river, which started from the Himalayan hill. The dam has constructed for the purpose of power generation, water supply, irrigation and flood control.

Three phase planning has been done for the power generation to get the full potential of the dam. The first phase (Phase I) was completed and starting to generate power in the year of 2006 with 1000 MW capacity. Phase two (2) of the Tehri power project was commissioned in 2011 at Koteshwar, which is downstream of the Tehri dam having a capacity of 400 MW. The third phase (Phase 3) with the planning of generating 1000 MW by using pumped storage plant is to be commissioned by 2021. Tehri dam is the main source of water supply for millions of people living near the reservoir. Irrigation canals and farms of Uttar Pradesh state depend upon water from Tehri reservoir for Rabi crop.

1.2.1 Tehri catchment

The total Tehri catchment area is 7295 km^2, out of which 2328 km^2 is snowbound and glaciers catchment. The maximum and minimum elevation of the Tehri catchment is about 7000 m and 600 m respectively from the above mean sea level. The catchment has seasonal snowline with descend in the eastern part of the Himalayas to an altitude of 3200 m and in the western part of the Himalayas to an altitude of 2600 m in March. The catchment is located between latitude 30˚ 20’ 20” N to 31˚ 27’ 30” N and Longitude 78˚ 09’ 15” E to 78˚ 28’ 54” E. The catchment also receives uneven rainfall distribution mostly from south-west monsoon and receives the light showers during the winter season. The long term average annual rainfall of Tehri catchment ranging from 1016 to 2630 mm. The catchment receives atmospheric temperature range from 2˚C to 40˚C during the winter season and summer season.

1.2.2 Tehri reservoir

The area of the reservoir at FRL (Full reservoir level) is 42 km^2. The length of the reservoir from the dam site embankment extends up 45 km upstream to Dharasu on river Bhagirathi and goes 25 km length to Ghansali on river Bhilangana. The reservoir has a gross and live storage capacity of 3549 MCM and 2616 MCM, respectively. The spillways of Tehri dam has been planned and designed for a maximum flow of 15,540 m^3/s.

1.3 OBJECTIVE OF THE STUDY

The purpose of the present study is to develop stochastic models for daily streamflow forecasting in Tehri catchment mainly into three sub-basins, i.e. Bhagirathi, Bhilangana and

Balganga, which are important tributaries of the Tehri reservoir. For all the tributaries gauge and discharge, data are being observed manually as well through the non-contact gauge and discharge sensors. As a first step, gauge and discharge data for all the locations were corrected and processed, and then the flow forecast models were developed. This objective resulted in the following sub-objectives:

i. Verification and preparation of the cross-section river profiles for the two sites, namely Ghansali and Sarasgaon.
ii. Development of a stage-discharge relationship (rating curve equations) for Ghansali and Sarasgaon sites using ANN and method of least squares.
iii. Development of stochastic models using AR, ARX, ARMA and ARMAX models for monsoon and non-monsoon daily flows for the three sub-basin (Bhagirathi, Bhilangana and Balganga River) of Tehri catchment.
iv. Checking the Goodness of fit for the selected model and evaluation of the model performance.
v. Selecting the best linear stochastic model for modelling daily streamflow forecasting in Bhagirathi, Bhilangana and Balganga River.
vi. Selection of best stochastic model using Akaike Information Criteria (AIC).
vii. Development of AR, ARX, ARMA and ARMAX models' programs using a computer technique (e.g. R-studio programming language).

1.4 THESIS OUTLINE

The research report is divided into eight main chapters, which have been discussed. Chapters as follows:

Chapter 1 presents an introduction for the present study, importance of the research study, brief details of the Tehri dam and its catchment, and objectives of the study.

The review of literature pertaing to stochastic models is presented in **Chapter 2**.

Chapter 3 provides an explanation of the study area and data used for the particular watershed area.

Chapter 4 presents the details of the methodology used for development of the rating curve equations and also the results of the analysis.

Chapter 5 describes the methodology and results of daily streamflow forecasting using stochastic models.

Chapter 6 presents the conclusions from the presented study and recommendations and scope for future work are drawn.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter presents the review of literature pertaining to the use of stochastic models for inflow forecasting in the next section.

2.2 STREAMFLOW FORECASTING USING Stochastic models

Stochastic models have been widely used for modelling of hydrological processes, which are primarily stochastic in nature. Hipel (1977), Kottegoda (1980), Salas et al. (1980), and Kumar (1983) discuss the theoretical aspects of time series modelling and stochastic models. In flood forecasting, some researchers have used the stochastic model's world over. In context of flood forecasting for Indian Rivers also stochastic models have been widely used (see e.g. Chander and Sapolia, 1976 for river Brahmaputra; Goel, 1982; and Goel and Chander, 1984 for flood stage forecasting using ARMAX models for river Marchur in Central India; Gosain and Chander, 1984 for river Yamuna etc.). Some of the recent applications (after 2010) of stochastic models for streamflow forecasting are reviewed in this section as follows:

Bogner and Pappenberger (2011) applied Autoregressive model with and without exogenous variable input (ARX and AR, respectively), as well as wavelet transforms (VARX), in a flood forecasting system at the Danube catchment.

Lohani et al. (2012) performed hydrological modelling and forecasting monthly reservoir inflow at Bhakra Dam using autoregressive (AR), artificial neural networks (ANNs) and adaptive neural-based fuzzy inference system (ANFIS). The results of the autoregressive (AR) model showed that the model could be useful for forecasting monthly reservoir inflow at Bhakra Dam.

Dutta et al. (2012) used several well-known TS based linear techniques and RR models for evaluation of streamflow forecasting in two sub-basins, namely upper Murray Basin and the Murray-Darling Basin in Australia. The model results showed that the ARMAX model provided better results for Bandiana station rather than the AR model for up to 3-days streamflow forecasting.

Sarhadi et al (2014) have applied the ARMA and ARMAX model to determine the daily and monthly snow water equivalent (SWE) in Ontario, Canada. The results showed that the ARMAX model performed better than the ARMA and SARIMA model for the forecasting of daily SWE.

Akouemo and Povinelli (2017) have performed Data Improving in Time Series by using autoregressive with and without exogenous variable inputs (ARX) and artificial neural network (ANN) models. Two approaches were applied for the detection and imputation of anomalies in time series data. The paper results demonstrated that the proposed approaches were able to identify and impute anomalous data points.

Ouyang et al. (2017) applied the Multi-Objective Genetic Algorithm (MOGA) to forecasting the models in his study by using stochastic models with ARX (Auto-Regressive model with exogenous variable inputs). In the study, they employed MOGA for the search for the optimal combination of non-sequential regressors in binary strings. The results of his study showed the optimal models performed good and have better the inundation forecasting in every time and time shift error and as well as error distribution.

Agrawal (2018) applied stochastic models for real-time inflow forecasting for three sites of Tehri catchment. Some of the conclusions drawn in the study are given below:

i. Autoregressive and Autoregressive models with exogenous inputs have performed very well for all the sites of Tehri catchment.
ii. For the forecasting of monsoon flows with 6 hours lead time ARX (1,1) model has performed very well with NSE more than 82% at Tehri dam.
iii. Reservoir levels were forecasted 78% of the time within the range of $\pm$ 10 cm accuracy in 6 hourly forecastings during monsoon season.
iv. In one- day advance forecasting during the non-monsoon season, 47 % of the forecasts are within the range of $\pm$ 5 cm accuracy without updating of parameters. With updating of parameters of the model, these models performed far better. During the period 18.11.2018 to 30.12.2018 more than 90% of the forecasts are within the range of $\pm$ 5 cm accuracy.
v. Stochastic models are easy to use and require fewer data. These models have performed very well in operation inflow forecasting system for Tehri dam.

Based on the literature review, the applicability of stochastic models has been explored further using extended data beyond Nov. 2018 and the results have been compared with HEC-HMS model.

CHAPTER 3

STUDY AREA

3.1 GENERAL

A detailed description of the study area in terms of its locale, precipitation, runoff, geology and soils and snow cover is presented in this chapter.

3.2 STUDY AREA

The study area for the present study is Tehri catchment (Fig. 3.1). Tehri catchment has a total area of 7295 km^2 (Table 3.1). Tehri catchment has two main rivers, namely Bhagirathi and Bhilangana. The Bhagirathi catchment is positioned between longitude of 78˚ 09’ 15” E to 79˚ 24’ 55” E and latitude 30˚ 20’ 20” N to 31˚ 27’ 30” N. The elevation difference of the catchment is very high, which ranges from 617m to 7000m above MSL. The Bhilangana catchment is positioned between longitude of 78° 38' 10.68" E to 79° 39' 24.48" E and latitude of 30° 25' 48.54" N to 30° 50' 36.708" N. The elevation difference of the catchment ranges from 840m to 3,717m above MSL (Figure 3.1). The main tributary of Bhilangana river is Balganga. The details of the catchment area are given in Table 3.1.

A multipurpose Tehri dam is constructed across river Bhagirathi nearly 1.5 km downstream of its confluence with river Bhilangana at Tehri in Uttarakhand. The live storage and gross storage of the Tehri dam are 3540 and 2615 MCM, respectively. The catchment mostly receives rainfall from the southwest monsoon, but the distribution is uneven. Also, it receives light showers during the winter months. The average annual rainfall of the catchment varies from 1016 to 2630 mm. The catchment has eleven (11) number of automatic weather stations, and six (6) number of automatic water level stations (Figure 3.1 and Table 3.2).

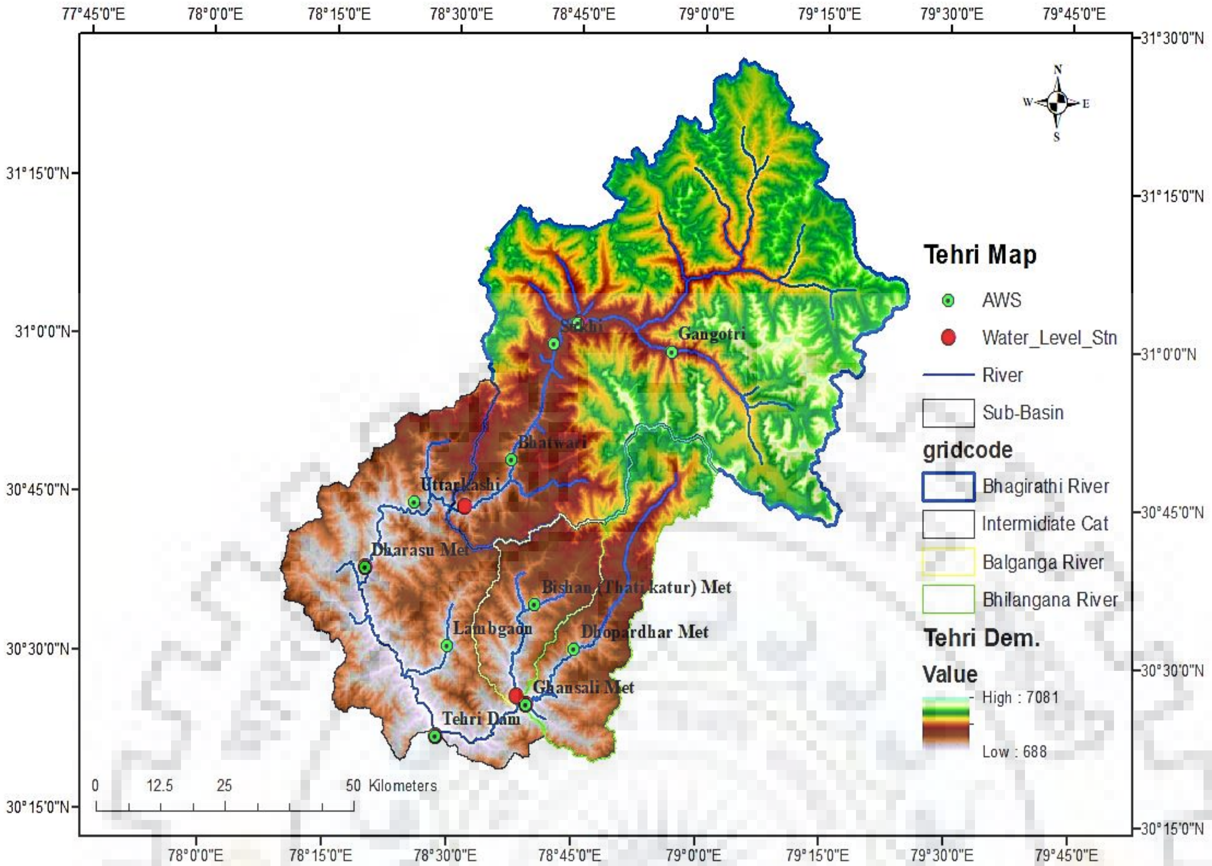


Figure 3.1.Tehri Catchments with study area catchments.

Table 3.1.Description of the study area and gauging stations

River	Station	Country	Drainage Area (km^2)	Elevation start (m)
Bhilangana	Ghansali	India	784.34	850
Balganga	Sarasgaon	India	486.43	860
Bhagirathi	Dharasu	India	4260.03	830
Intermediate catchment surrounding Tehri Reservoir		India	1764.02	617

Table 3.2. Details of Hydro-Meteorological stations of Tehri catchment

S. No.	Locations	Latitude	Longitude	Remarks
1	Gangotri	30°59’40.89” N	78°56‘13.00” E	AWS with snow gauge sensor
2	Harshil	31°02’07.72” N	78°45’04.02” E	AWS with snow gauge sensors; Automatic and manual G&D stations
3	Sukkhi	30°56’46.14” N	78°41’15.74” E	AWS
4	Bhatwari	30°49’06.92” N	78°37'05.17” E	AWS
5	Uttarkashi	30°43’42.80” N	78°25’25.53” E	AWS; water level recorder, Manual meteorological observatory
6	Dharasu	30°38'28.61” N	78°19’45.59” E	AWS; Automatic and manual G&D stations
7	Lambgaon	30° 37' 48” N	78° 33' 10” E	AWS
8	Tehri	30°22’46.16" N	78°28’59.29" E	AWS; water level recorder Manual observatory
9	Dhopardhar	30° 21’ 39” N	78° 47’ 24" E	AWS
10	Ghansali	30°25’46.57" N	78°39’23.71" E	AWS; Automatic and manual G&D stations; Manual meteorological observatory
11	Sarasgaon	30^{0}26’37.13” N	78^{0}38’10.90” E	Automatic and manual G&D stations
12	Thati Kathur (Bishan)	30°34’43.31” N	78°38’55.81” E	AWS

3.3 PRECIPITATION

Mostly Himalayan catchments in Indian Northside receive heavy rainfall through the south-west monsoon season which extends starts from June to September end. The climate of the catchment varies with different elevations and aspect changes, and the area is cold generally.

3.4 RUNOFF

The runoff of the Tehri catchment depends on two sources, viz. the snowmelt which occurs from the snow-covered areas and glaciers in uphills, and rainfall in the lower catchment. The contribution of snow melt and base flow makes the river as a perennial. The amount of snowmelt and extent of snow-covered area in the catchment vary from year to year and also within the year.

3.5 GEOLOGY AND SOILS.

The rock at dam site consists of the Chandpur Phyllite. Based on lithological characteristics and engineering properties, this has been classified into broadly three grades viz. Grade I (Phyllite Quartize), Grade II (Quartzitic Phillite) and Grade III (Schistose Phyllite). Riverbed consists of large boulders. Average upstream slope of the river is 1: 22.

3.6 SNOW COVER

In the present study, the MODIS dataset has been used to find the snow covered area for the catchment. The results show that mostly 50 % of the catchment is snow-covered in winter and most part of this is temporary snow, which melts during the summer season. The permanent snow line is located approximately above 4500 m (Agrawal, 2009).

CHAPTER 4

DEVELOPMENT OF RATING CURVES

4.1 GENERAL

A rating curve or stage-discharge curve is a plot between stage and discharge at the gauging location of a stream. The development of the stage-discharge relationship in a river is important as it gives the estimate of discharge corresponding to a stage without direct measurement of discharge. In the past, a number of researchers have developed the stage-discharge relationship using different techniques like least squares method, and ANN etc. (Goel, 2011; Mir & Dubeau, 2014).

In the present study, rating curves have been developed for Bhilangana at Ghansali and Balganga at Sarasgaon using method of least squares and artificial neural network (ANN) and their performance has been evaluated.

4.2 DATA USED AND PRELIMINARY ANALYSIS

The daily stage and discharge data from June 2016 to November 2018 of two gauging stations, namely Ghanshali (Bhilangana river) and Sarasgaon (Balganaga river) have been used for the present study. The datasets were divided in two parts for model calibration and validation purpose. The data from June 2016 to December 2017 were used for calibration and data from January 2018 to November 2018 were used for validation.

The preliminary analysis of the data was started by grouping the observed daily discharge data and corresponding stages. The stage-discharge graphs are plotted for the two sites. The outliers are removed after plotting the data. The details of total number of data sets used for developing the stage-discharge relationship by least squares method and the outliers are given in Table 4.1.

The data scrutiny and analysis included the following steps:

- Screening of data series and removal of outliers
- Graphical comparison of streamflow and rainfall data

Table 4.1.Observed discharge and stage data summary used for developing stage-discharge relationship by least square method

S. No	Stations	Data Type	Unit	Total No.	Outlier
1.	Bhilangna	Discharge_Stage	m^3/s_m	1100	62
2.	Balganga	Discharge_Stage	m^3/s_m	1000	114

The summary of the dataset used for calibration and validation using ANN method is given in Table 4.2.

Table 4.2. Summary of the observed discharge and stage data used for developing a stage-discharge relationship by ANN method.

S. No	Stations	Data Type	Unit	Total No.	Outlier	Total
1.	Bhilangna	Discharge_Stage	m^3/s_m	700	300	1000
2.	Balganga	Discharge_Stage	m^3/s_m	560	240	800

4.2.1 River cross sections and plots of stage discharge curves for the two sites

The cross sections of the sites under consideration are plotted Fig. 4.1 and 4.2.

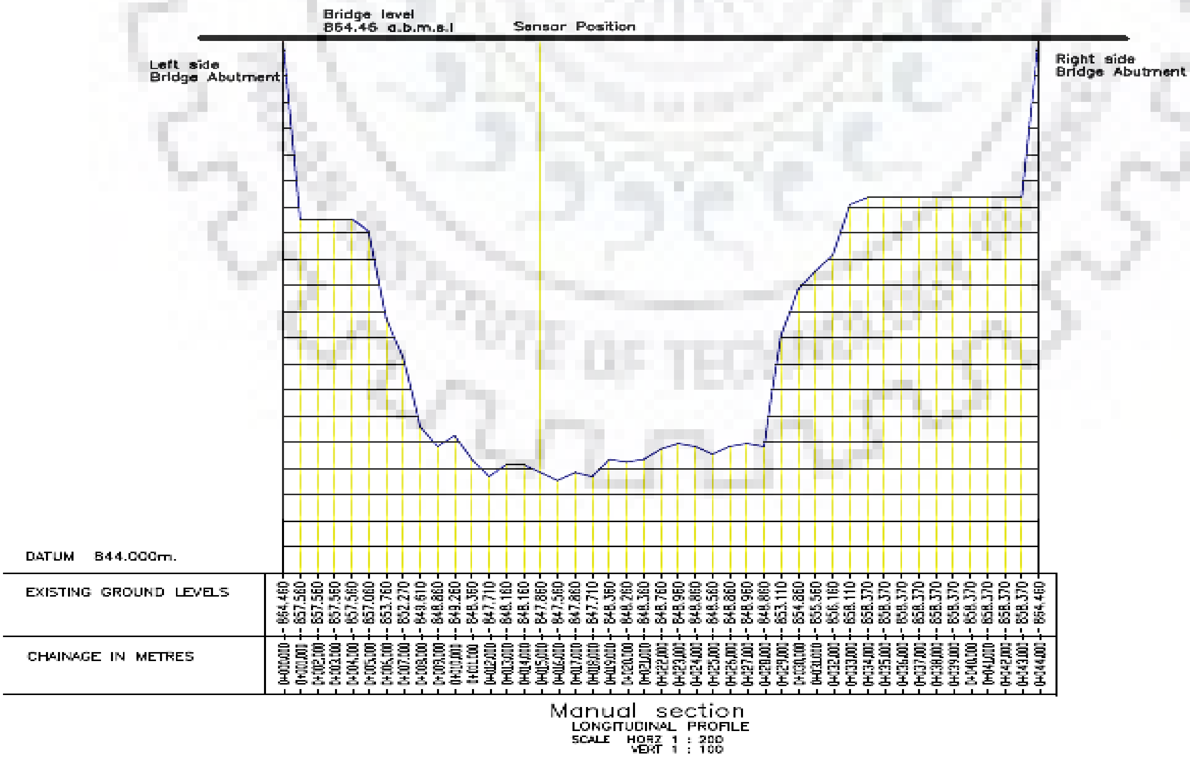


Figure 4.1.Cross-section profile at Bhilangana catchment

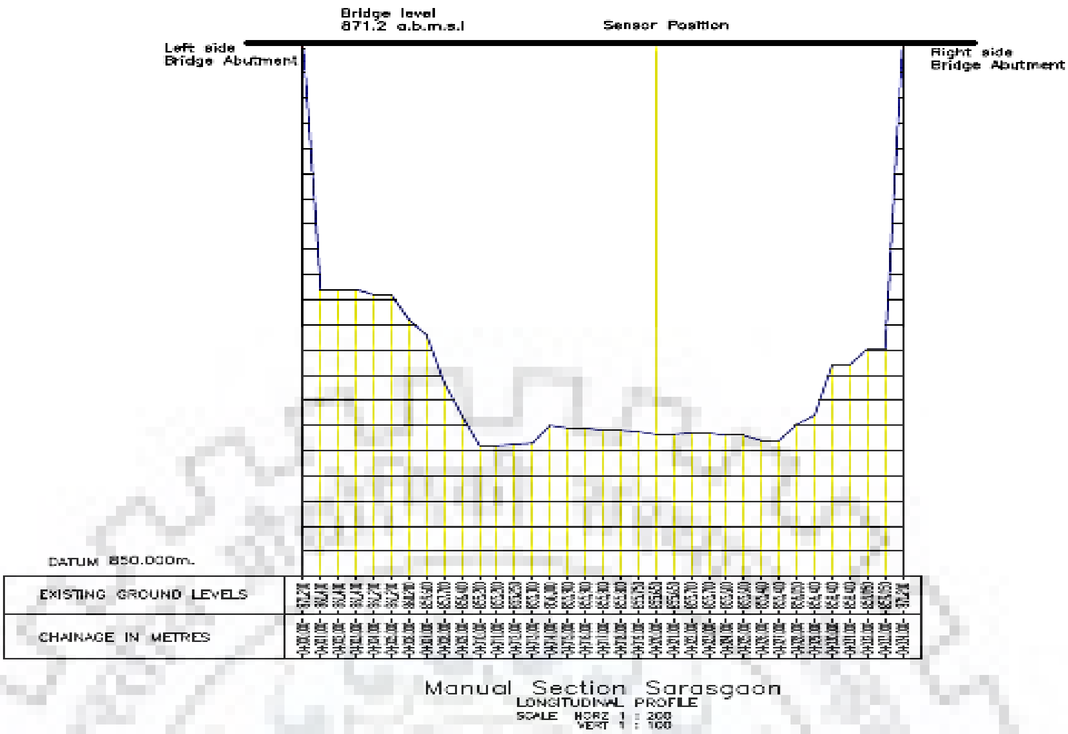


Figure 4.2. Cross-section profile at Bhilangana catchment

The arithmetic plots for the two sites are given Fig. 4.3 and Fig. 4.4. The log-log plots are given in Fig. 4.5 and 4.6.

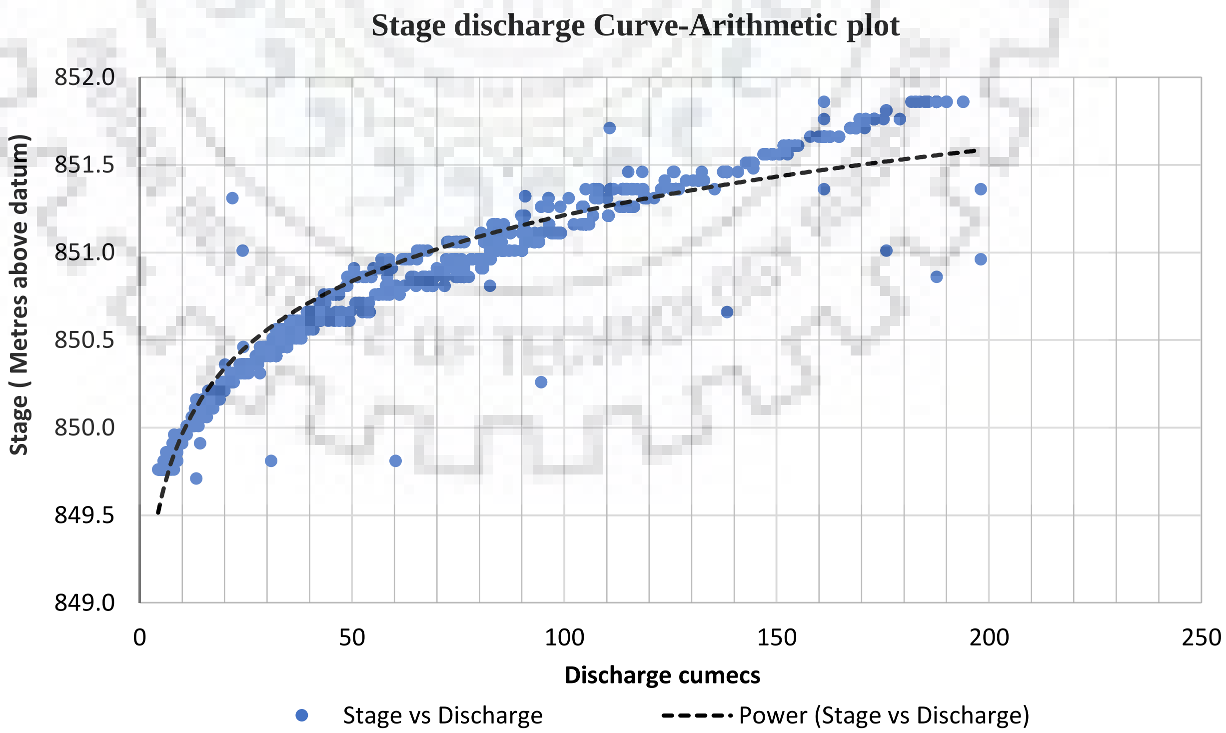


Figure 4.3. Arithmetic plot stage discharge for Bhilangana river at Ghansali

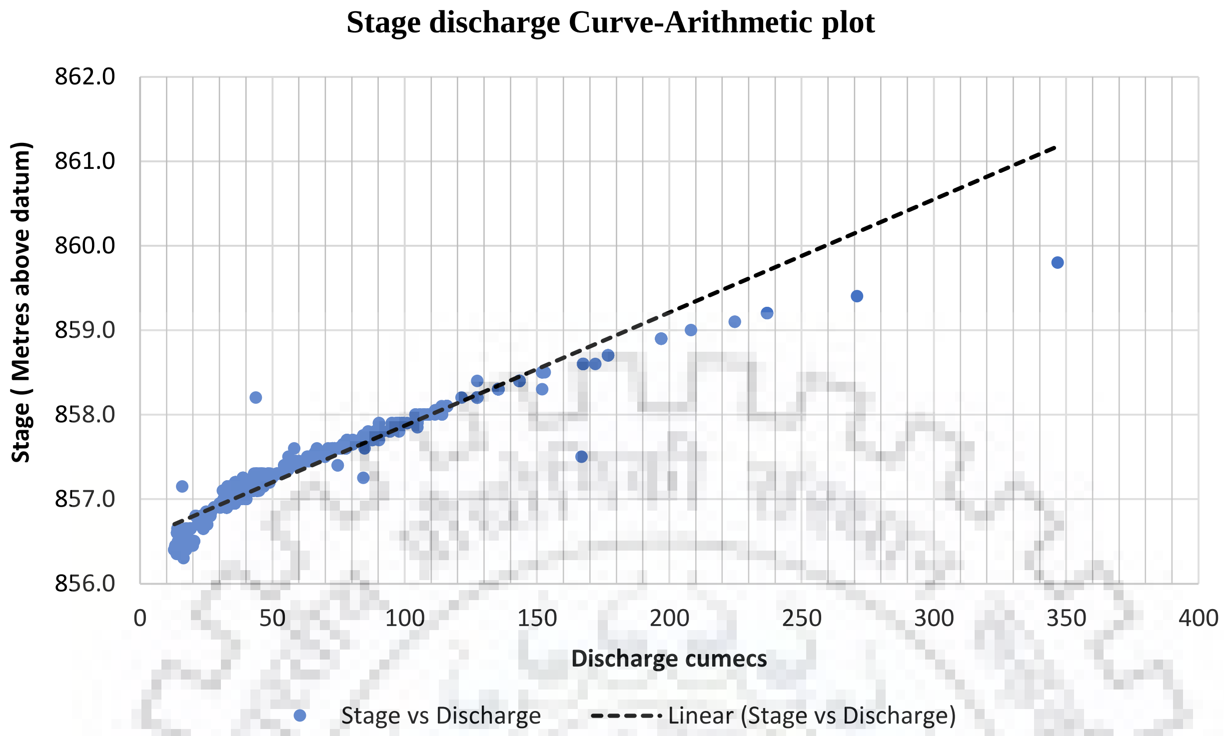


Figure 4.4. Arithmetic plot stage discharge for Balganga river at Sarasgaon

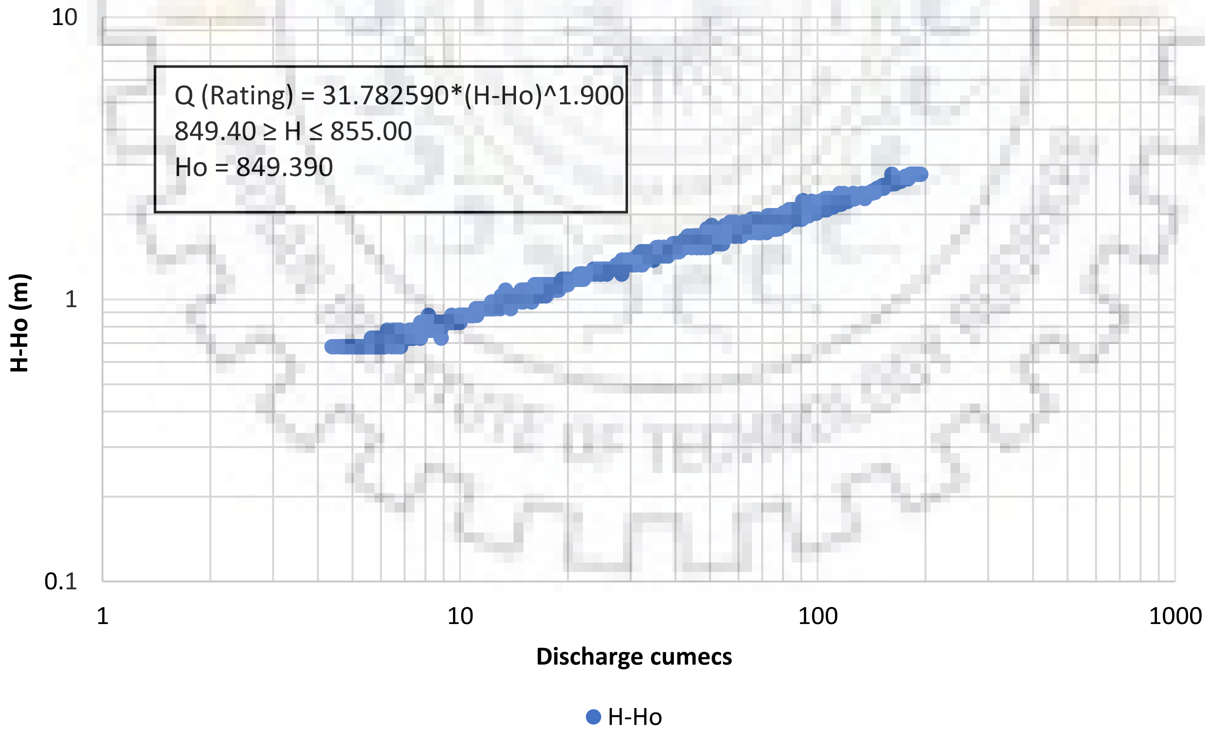


Figure 4.5. Log-Log plot for Bhilangana at Ghansali

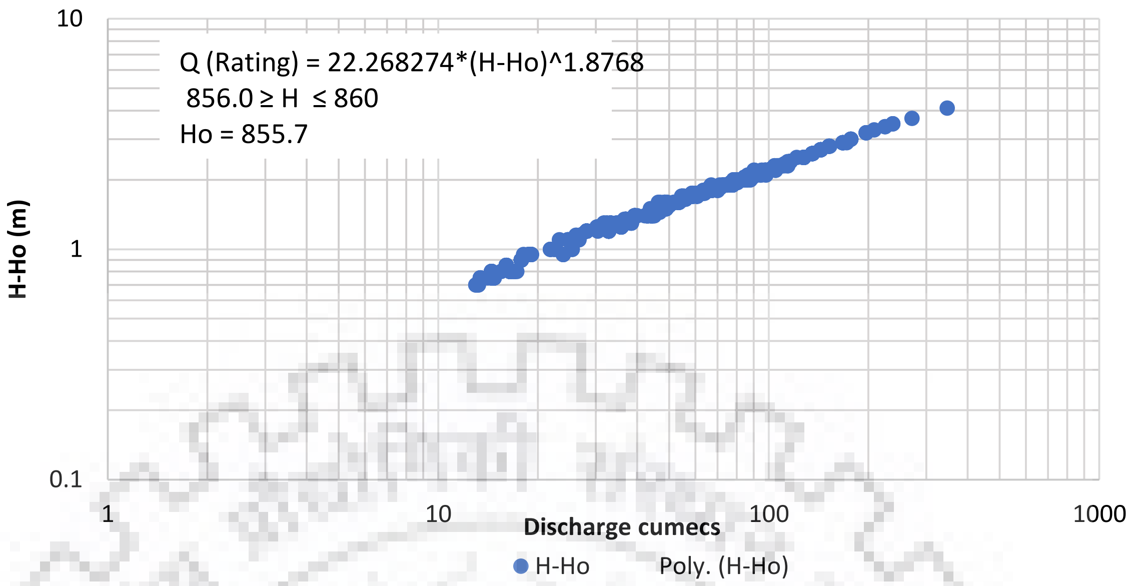


Figure 4.6. Log-Log plot for Balganga at Sarasgaon.

4.3 RATING CURVE DEVELOPMENT USING METHOD OF LEAST SQUARES

The following equation 4.1 to equation 4.5 are used to develop the relationship between stage and discharge using method of least squares:

$$Q = \mathrm{a}(\mathrm{H} - \mathrm{Ho})^b \tag{4.1}$$

$$\mathrm{Log}_e\mathrm{Q} = \mathrm{Log}_e\mathrm{a} + \mathrm{bLog}_e(\mathrm{H}\text{-}\mathrm{H}_o) \tag{4.2}$$

Or $$\mathrm{Y} = \mathrm{AX} + \mathrm{B} \tag{4.3}$$

Where,

Y=log$Q$; A=n; X=log(H−Ho); B=log$K$.

Using regression analysis, the values of A and B can be calculated through the following relations

$$A = \frac{\mathrm{N}\sum(\mathrm{XY}) - (\sum \mathrm{X})(\sum \mathrm{Y})}{\mathrm{N}(\sum \mathrm{X}^2) - (\sum \mathrm{X})^2} \tag{4.4}$$

$$B = \frac{\sum \mathrm{Y} - A(\sum \mathrm{X})}{N} \tag{4.5}$$

Where, Q = Stream discharge; H = Stage height; Ho = a constant representing the gauge reading corresponding to zero discharge; a and b are the rating curve constants.

The values of '*a*' and '*b*' from physical consideration of the cross section are given by the following equations;

$$a = (1/n)\ WS^{1/2} \tag{4.6}$$

$$n = 0.034d^{1/6} \tag{4.7}$$

Where, *W* is the top width of the channel, *S* is the bed slope and *n* is Manning's coefficient, *d* is medium size of the bed materials in mm. The typical value of '*n*' for natural rivers are (Henderson (1966)):

The clean and straight river channel 0.025 – 0.03

Winding with pools and shoals 0.033 – 0.04

Very weedy, winding and overgrown 0.075 – 0.15

The values of *b* for different types of cross sections are given below:

- For rectangular shape: 1.6
- For triangular shape: 2.5
- For parabolic shape: 2
- For irregular shape: 1.6 to 1.9

The values of '*a*' and '*b*' in equation 4.1 should be cross verified with the values of '*a*' and '*b*' from physical considerations.

4.4 ARTIFICIAL NEURAL NETWORK (ANN)

An Artificial Neural Network (ANNs) is a system based on the operation of biological neural networks. The concept of ANN was developed in 1943 by Warren McCulloch and Walter Pitts, who suggested the conceptualization of human brain function built on a network of interconnected cells. However, the use of ANN in hydrological applications started in 1990s and picked up momentum from 2000 after the publication of ASCE task committee report on Application of Artificial Neural network in Hydrology in ASCE Journal of Hydrologic Engineering. The architecture of Neural Network is designed in three layers, called the input layer, hidden layer(s), and output layer. In recent years numerous types of artificial neural network (ANN) have been developed, such as Feedforward, radial basis, recurrent and multilayer perceptron neural network. Gallant (1990) reported that the multilayer perceptron (MLP) and the feedforward ANN is the most commonly used type of ANN. In the present study, the multilayer perceptron (MLP) based approaches have been used to develop the stage-discharge relationships for two sub-basins of Tehri catchment.

A multilayer perceptron (MLP) is a class of feedforward artificial neural network, which consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function.

The ANN uses the following expression:

$$y = f\left(\sum_{i=1}^{n} x_i w_i + b\right) \tag{4.8}$$

where, the input are X subscripts (i = 1,2, ... n) and corresponding model weights are w with subscripts (i = 1,2, … n), b is the bias, y is the output and f(.) is the activation function. The actual node input expressed as,

$$\text{net} = \left(\sum_{i=1}^{n} x_i w_i\right) \tag{4.9}$$

4.4.1 Learning in multilayer feed-forward networks

The weight matrix for the training is the initial step for developing an ANN model. There are two types of learning mechanisms or training, i.e. supervised and unsupervised. In the present study, supervised learning is used for training the dataset. There are a number of algorithms like backpropagation, Conjugate gradient algorithm, radial basis function, cascade correlation algorithm etc. available for supervised learning. In the present study, a backpropagation algorithm was used for the training of a multilayer perceptron (MLP).

4.4.2 Back-propagation (BP)

In back-propagation, the minimization of errors for the target and calculated (simulated) output is done by the modification of the network weights. Usually, the algorithm is designed based on the correction of the error. Backpropagation algorithm includes the two phases, i.e. Forward phase and Backward phase. All the parameters weight normally initialized and updated (in each iteration) by using back-propagation and feed-forward method.

I. Feed-forward calculation

For Feed-forward calculation, the input nodes in the layer give the signal input to the hidden layer and at the same time to the output layer as follows:

The j^{th} node for net input in the hidden layers is expressed by

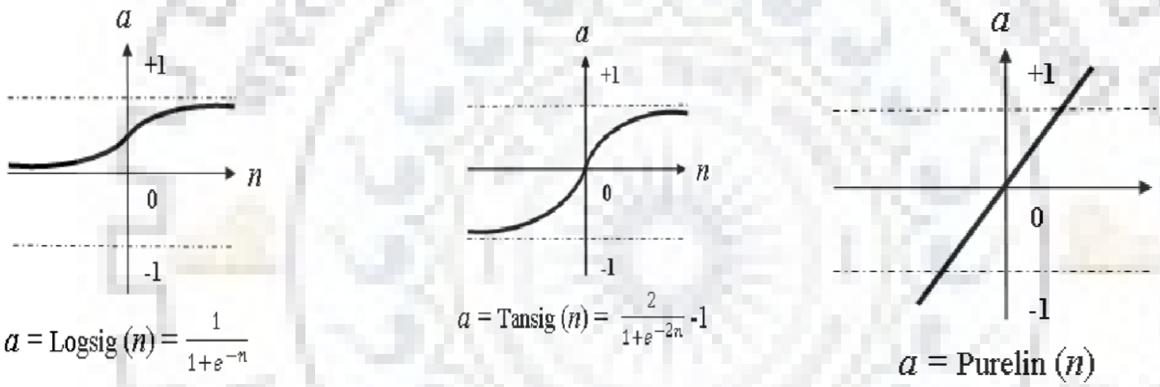
$$neth_j = \left(\sum_{i=1}^{n_i} wh_{ji} x_i \right) \tag{4.10}$$

Where, n_i, number of neurons (in the input layer) and the connection weight is wh_{ji}, i^{th} represent the node input layer and j^{th} node hidden layer. The following expression is the output of the j^{th} node hidden layer h_j which known as,

$$h_j = f\left(neth_j\right) \tag{4.11}$$

Where, $f(.)$ representing the sigmoid or activation function.

$$h_j = \frac{1}{1 + exp\left(-neth_j\right)} \tag{4.12}$$



(a) Log-sigmoid (b) Tan-sigmoid (c) Purelin

Figure 4.7. Activation of Transfer functions and their range

The net input is equal to the k^{th} node output layer is specified by the following relationship.

$$neth_k = \left(\sum_{j=1}^{n_h} wo_{kj} x_j \right) \tag{4.13}$$

Where, n_h representing the number of neurons (hidden layer) and wo_{kj} is a connection weight for nodes and j^{th} hidden layer as well as k^{th} representing a node output layer.

$$y_k = f\ (nety_k) \tag{4.14}$$

Now operating through the pure linear activation function, it can be written as,

$$y_k = \text{purelin}\ (nety_k) \tag{4.15}$$

II. Error back-propagation

The hidden layers and Input layers are propagated back by the error from the computed output layer to determine the updates for the weights. The technique is derivative from the known gradient descent process where the weights modernizing is done by moving the negative gradient alongside the multidimensional surface of the error function. The following expression bellow is the mean sum of square error E.

$$E = \frac{1}{2N}\sum_{k=1}^{n_o}(y_k - t_k)^2 \tag{4.16}$$

Where, t_k representing the desired output (target k[th]) and y_k is the computed output for the same node.

4.4.3 Model Development

For Artificial Neural Network model, a different combination of TS with three antecedent gauge and discharge values were developed for the analysis as follows:

Model one: Q_t = f(H_t'),

Model two: Q_t = f (H_t', H_{t-1}'),

Model three: Q_t = f (H_t', Q_{t-1}),

Where, Q_t is a discharge or river flow in cumecs and H_t represents river stage in meter all correspond with time t (*Ht-Ho)*.

The feedforward back-propagation tool of MATLAB version 9.2 has been used in the study for developing the relationship between the stage and discharge. Three input-output data sets were given into the ANN tools for model development. The input layer of the model is the river stage and discharge data set, and the output layer neuron is only discharged dataset. The '*nntools*' function of MATLAB was used for the development of the ANN model.

The gradient descent method was used for the adjustment of the weights and biases of the network during the simulation time. The adjustment of weight and bias was adjusted by using the learning function. The mean square error is the error performance function for the feedforward network (where the computation of the square error was done between the outputs network and target outputs). To obtain consistency in the results, the number of trials must be made randomly in order to resolve the uncertainties of the initial weights and stopping criteria

(Sahoo & Ray, 2006). The improvement of the developed model was frequently checked by testing data on iteration to avoid the overtraining.

4.4.4 Preprocessing of input data

Preprocessing of input data is essential for the adeptness of the training algorithm. MLP is very sensitive and can only be used by scaling the data. Initially, the input variables were standardized or rescaled to make sure that the datasets get the same attention during the training process. In this study, normalization of raw data series was applied to both stages and discharge data series.

4.4.4.1 Normalization of raw data series

The hyperbolic tangent (tansig) function was used in this study, in which the data sets were differentiated and monotonically increasing. The output of this function is in the range from -1 to +1, which is permanently bonded and then the input to the function may vary between -∞ to ∞. Another method is to rescale the data sets values with a mean of 0 and unit standard deviation (referred to as normalization). Here below is the expression of the normalization method.

$$x_n = \frac{x - \overline{x}}{x_{SD}} \tag{4.17}$$

The normalized data were then de-normalized using the following relationship.

$$x = (x_n \times x_{SD}) + \overline{x} \tag{4.18}$$

Where, *n* Number of observations in a data series; *x* is the data set, x_{SD} standard deviation, $\overline{x}$ is the mean of the data set.

4.5 MODEL PERFORMANCE EVALUATION

The performance of a model can be assessed by using different performance indicators. In the present study, three different performance indicators namely correlation coefficient (r), the root means square error (RMSE) and Nash Sutcliffe efficiency (NSE) (Nash & Sutcliffe, 1970) were used as per following equations:

$$\text{The coefficient of Correlation} = \frac{\sum_{i=1}^{N}(Q_O - \overline{Q}_O)(Q_p - \overline{Q}_p)}{\sqrt{\sum_{i=1}^{N}(Q_O - \overline{Q}_O)^2 \sum_{i=1}^{N}(Q_p - \overline{Q}_p)^2}} \tag{4.19}$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum_{i=1}^{N}(Q_{Oi}-\bar{Q}_{pi})^2}{N}} \tag{4.20}$$

$$\text{Nash Sutcliffe Efficiency} = 100 \times \left[1 - \frac{\sum_{i=1}^{N}(Q_O-\bar{Q}_p)^2}{\sum_{i=1}^{N}(Q_O-\bar{Q}_O)^2}\right] \tag{4.21}$$

Where, N is the number of observed data; Q_o is observed river flow, Q_p is predicted river flow, $\bar{Q}_O$ is mean observed river flow and $\bar{Q}_p$ is mean predicted river flow.

4.6 RESULTS AND DISCUSSION

The results obtained from both the least squares method and Artificial Neural Network (ANN) for developing the rating curve of Bhilangana river at Ghansali and Balganga river at Sarasgaon are discussed in this section.

4.6.1 Least Square Method

A higher degree of polynomial fit and best fits was obtained. The results of the calibration and validation process with and without outliers obtained for Bhilangana sub-basin are given in Table 4.3. For Bhilangana sub-basin, the model performance indicator R2, RMSE and NSE are 0.994, 9.253 and 99.066%, respectively during the calibration process without outliers, while during the validation process, the model performance indicator R2, RMSE and NSE are 0.986, 9.252 and 94.360%, respectively. The results of the calibration and validation process with outliers and without outliers obtained for Balganga sub-basin is given in Table 4.4. For Balganga river the model performance indicator R2, RMSE and NSE are 0.997, 2.891 and 99.276%, respectively during the calibration process without outliers, while during the validation process, the model performance indicator R2, RMSE and NSE are 0.938, 6.988 and 87.433% respectively. The equation used for calculating the stream flow is given in Table 4.5 for without outliers and in Table 4.6 for with outliers. The table represented the value of zero flow and the value of constant parameters of a and b. Figure 4.8 to 4.11 shows the graph plots of the relationship between stage and discharge with and without outliers, which will be used for calculating/computing discharge for the future.

Table 4.3. Representing performance and error results for Bhilangana River at Ghansali

Artificial Neural Network (ANN)	Number of Hidden Layer	Data with Outliers					
		Calibration			Validation		
		Coefficient of Correlation R^2	RMSE (m^3/sec)	Efficiency N %	Coefficient of Correlation R^2	RMSE (m^3/sec)	Efficiency N %
$Q_t = f(H_t)$	3	0.998	5.651	99.660	0.993	4.536	98.599
Method of least squares		0.989	13.470	98.133	0.986	9.252	94.360
Artificial Neural Network (ANN)	**Number of Hidden Layer**	**Data without Outliers**					
		Calibration			**Validation**		
		Coefficient of Correlation R^2	**RMSE (m^3/sec)**	**Efficiency N %**	**Coefficient of Correlation R^2**	**RMSE (m^3/sec)**	**Efficiency N %**
$Q_t = f(H_t)$	3	0.999	0.947	99.990	0.996	3.174	99.322
Method of least squares		0.994	9.253	99.066	0.986	8.424	95.324

Table 4.4. Representing performance and error results for Balganga River at Sarasgaon

Artificial Neural Network (ANN)	Number of Hidden Layer	Data with Outliers					
		Calibration			Validation		
		Coefficient of Correlation R^2	RMSE (m^3/sec)	Efficiency N %	Coefficient of Correlation R^2	RMSE (m^3/sec)	Efficiency N %
$Q_t = f(H_t)$	3	0.994	0.886	99.953	0.99	0.989	99.748
Method of least squares		0.989	6.399	97.446	0.938	6.937	87.616
Artificial Neural Network (ANN)	**Number of Hidden Layer**	**Data without Outliers**					
		Calibration			**Validation**		
		Coefficient of Correlation R^2	**RMSE (m^3/sec)**	**Efficiency N %**	**Coefficient of Correlation R^2**	**RMSE (m^3/sec)**	**Efficiency N %**
$Q_t = f(H_t)$	3	0.999	0.031	99.992	0.999	0.204	99.989
Method of least squares		0.997	2.891	99.276	0.938	6.988	87.433

Table 4.5. Summary of actual rating curve without outliers

S. No	Stations	Ho	b	a	Best fit	Rating Curve equation	Range of applicability
1.	Bhilangana	849.39	1.900	31.783	98.97	$Q = 31.783*(H\text{-}Ho)^{1.900}$	$849.40 \geq H \leq 855.00$
2.	Balganga	855.8	1.742	22.7743	99.58	$Q = 22.774*(H\text{-}Ho)^{1.742}$	$856.0 \geq H \leq 860$

Table 4.6.Summary of actual rating curve with outliers

S. No	Stations	Ho	b	a	Best fit	Rating Curve equation	Range of applicability
1.	Bhilangana	849.39	1.896	32.882	95.45	$Q = 32.882*(H\text{-}Ho)^{1.896}$	$849.40 \geq H \leq 855.00$
2.	Balganga	855.8	1.742	26.994	97.85	$Q = 26.994*(H\text{-}Ho)^{1.711}$	$856.0 \geq H \leq 860$

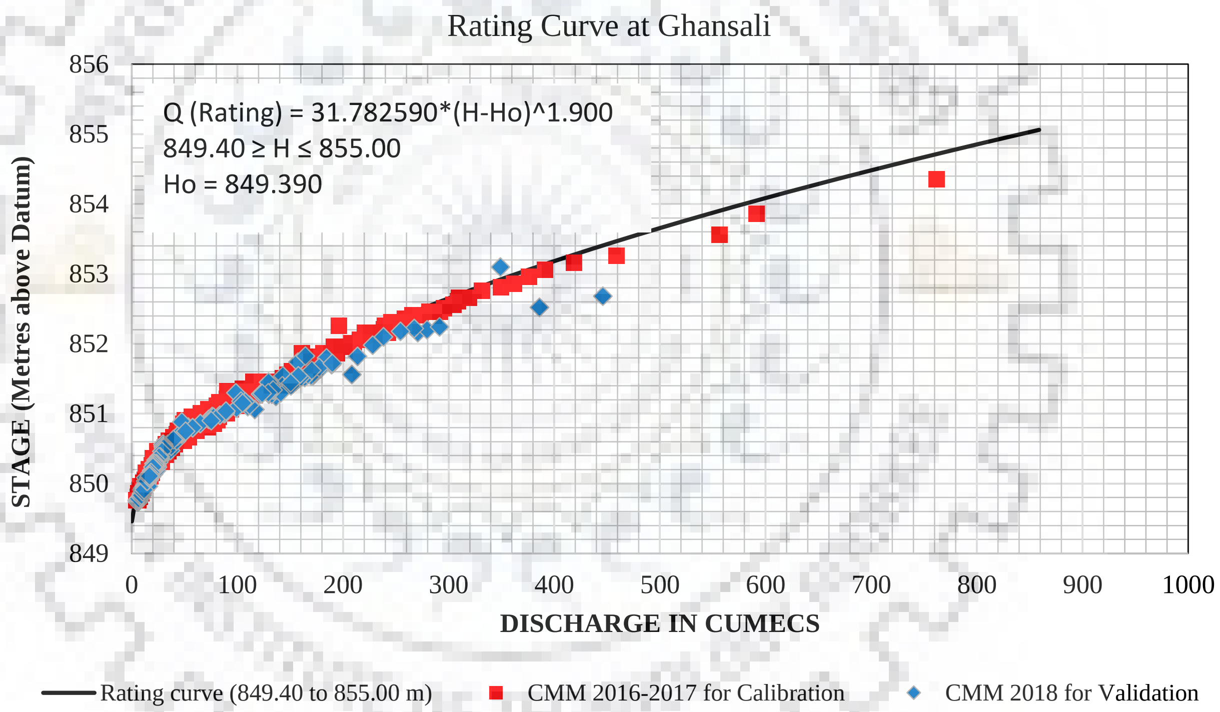


Figure 4.8.Rating Curve and Equation at Bhilangana River at Ghansali without outliers

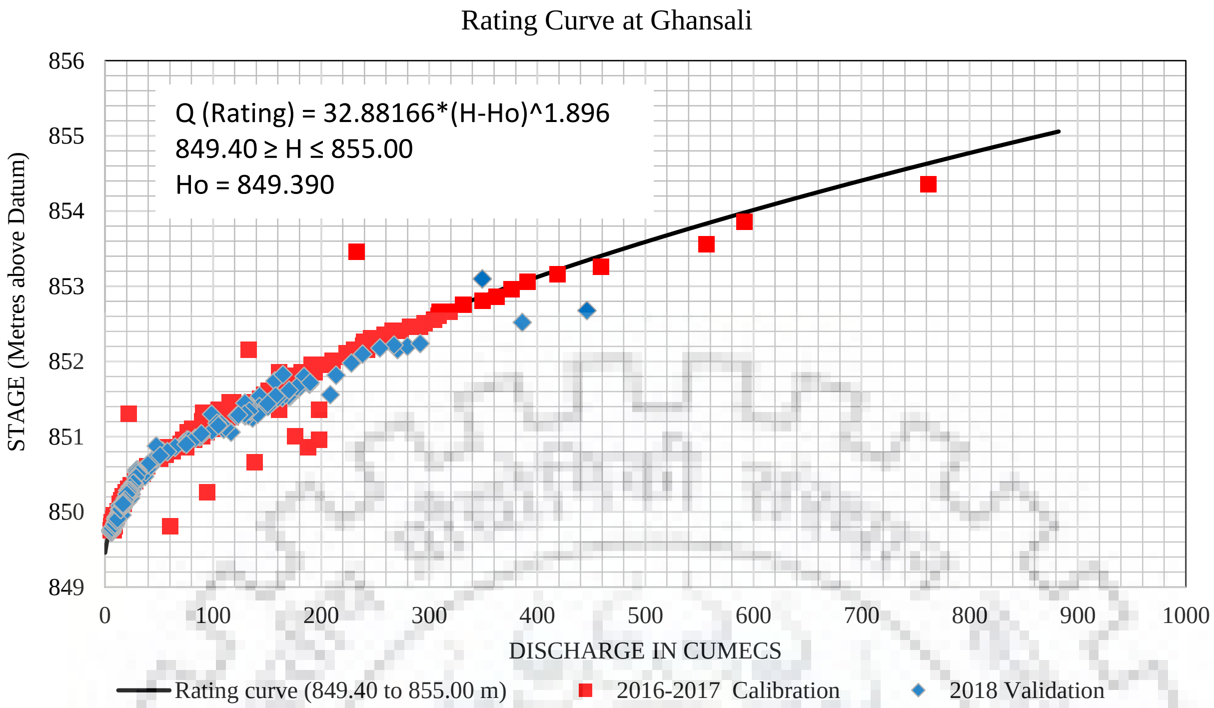


Figure 4.9. Rating Curve and Equation at Bhilangana River at Ghansali with outliers

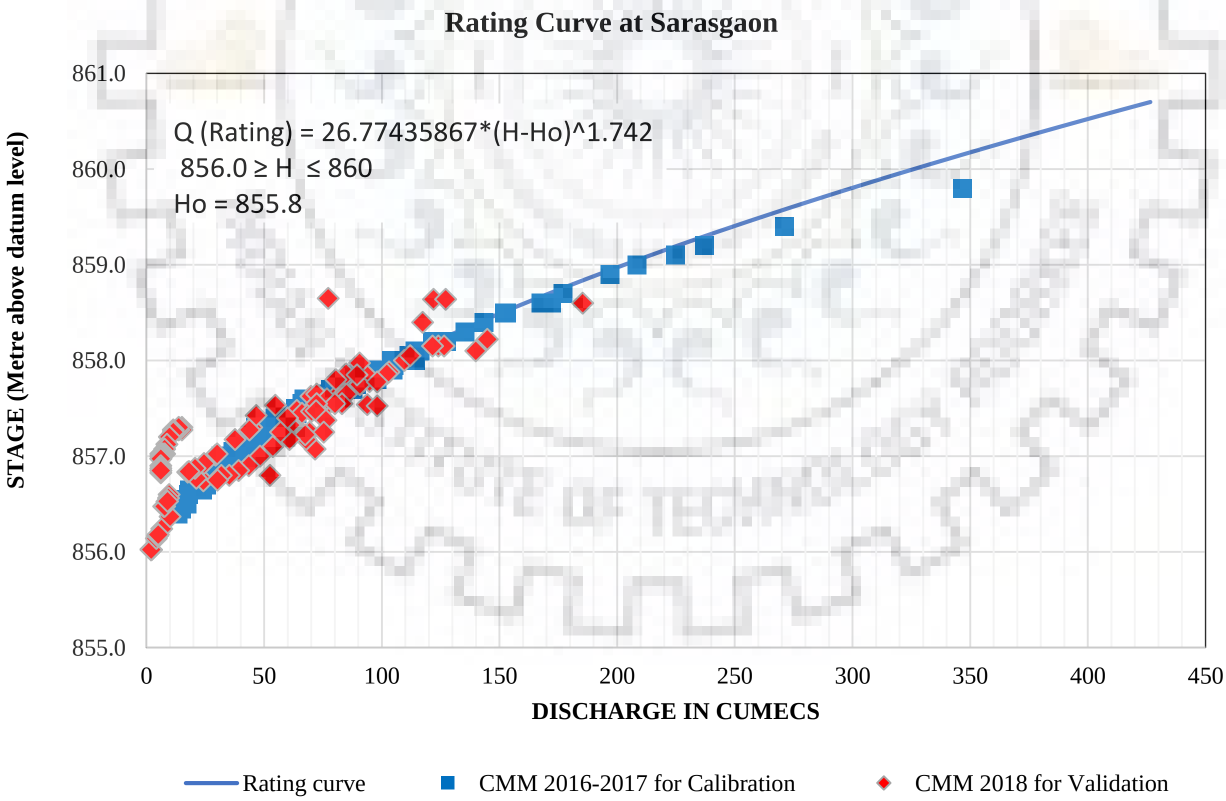


Figure 4.10. Rating Curve and Equation at Balganga River at Sarasgaon without outliers

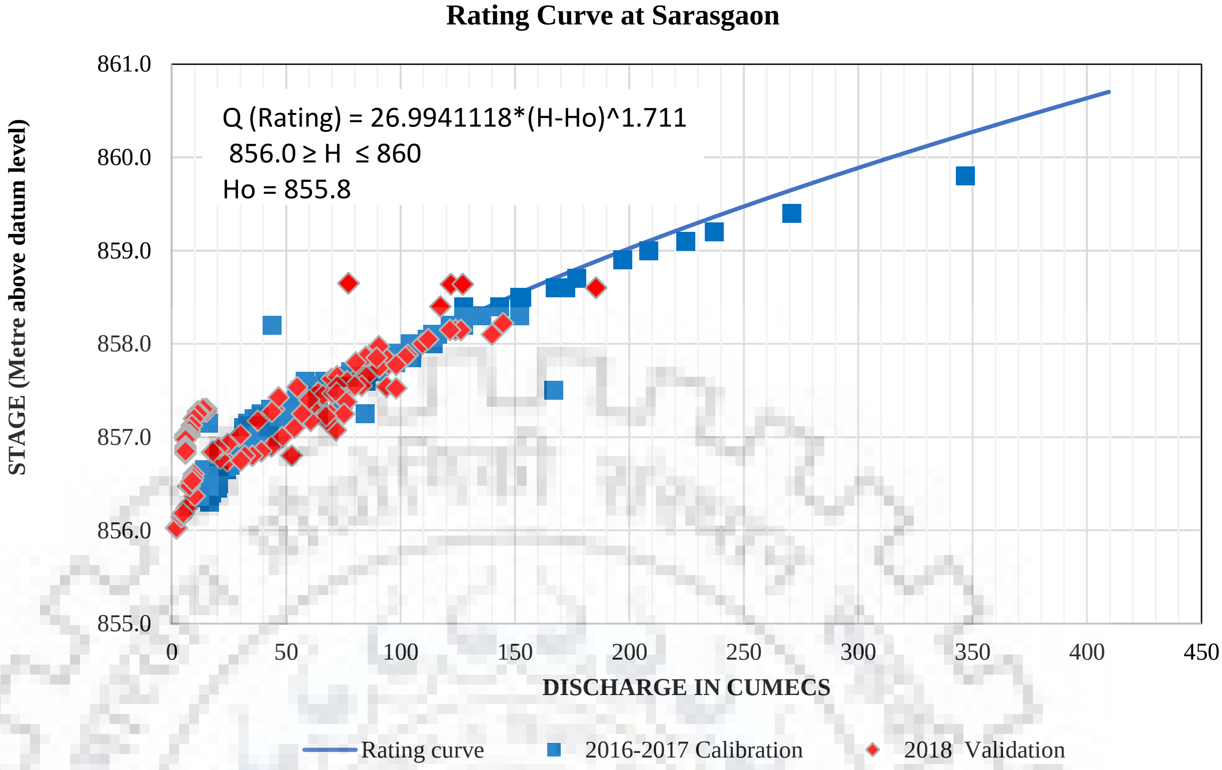


Figure 4.11. Rating Curve and Equation at Balganga River at Sarasgaon with outliers

4.6.2 ANN-Artificial Neural Network

Three number of the model combination with and without outliers were considered for the development of the stage-discharge relationship using ANN for two sub-catchments of Tehri catchment. The ANN results were evaluated by considering the model statistical analysis performance.

All the model combination is checked and the results are satisfactory for all the models. However, for Bhilangana sub-basin, the results of Model 1 without outliers and with 3 numbers of hidden layers are better as compared to other models results. Therefore, Model 1 is chosen during the calibration for the Bhilangana sub-basin. The results of the calibration and validation process for the Bhilangana sub-basin are given in Table 4.3. The calibration results of model 1 give the value of the model performance indicator R^2, RMSE and NSE as 0.999, 0.947 and 99.99 %, respectively. The validation results of the model give the value of the model performance indicator R^2, RMSE and NSE as 0.996, 3.174 and 99.322%, respectively.

For the Balganga river, the results of Model 1 without outliers and with 3 numbers of hidden layers are better as compared to the other model results. Therefore, Model 1 is chosen for the Balganga river. The results of the calibration process give the coefficient of correlation

R^2 as 0.999, RMSE as 0.031 and NSE as 99.992%. The results during the validation process give the coefficient of correlation, R^2 as 0.999, RMSE as 0.204 and NSE as 99.989%. The results are also shown in graphs. Figure 4.12 and 4.13 shows the results of calibration (training) and validation of the chosen model for Bhilangana and Balganga sub-basins, respectively. Figure 4.14 to Figure 4.17 representing the comparisons of the observed and computed streamflow with and without outliers for two different methods used in simulating the streamflow, the figures show the results with above and below 10% of the observed streamflow.

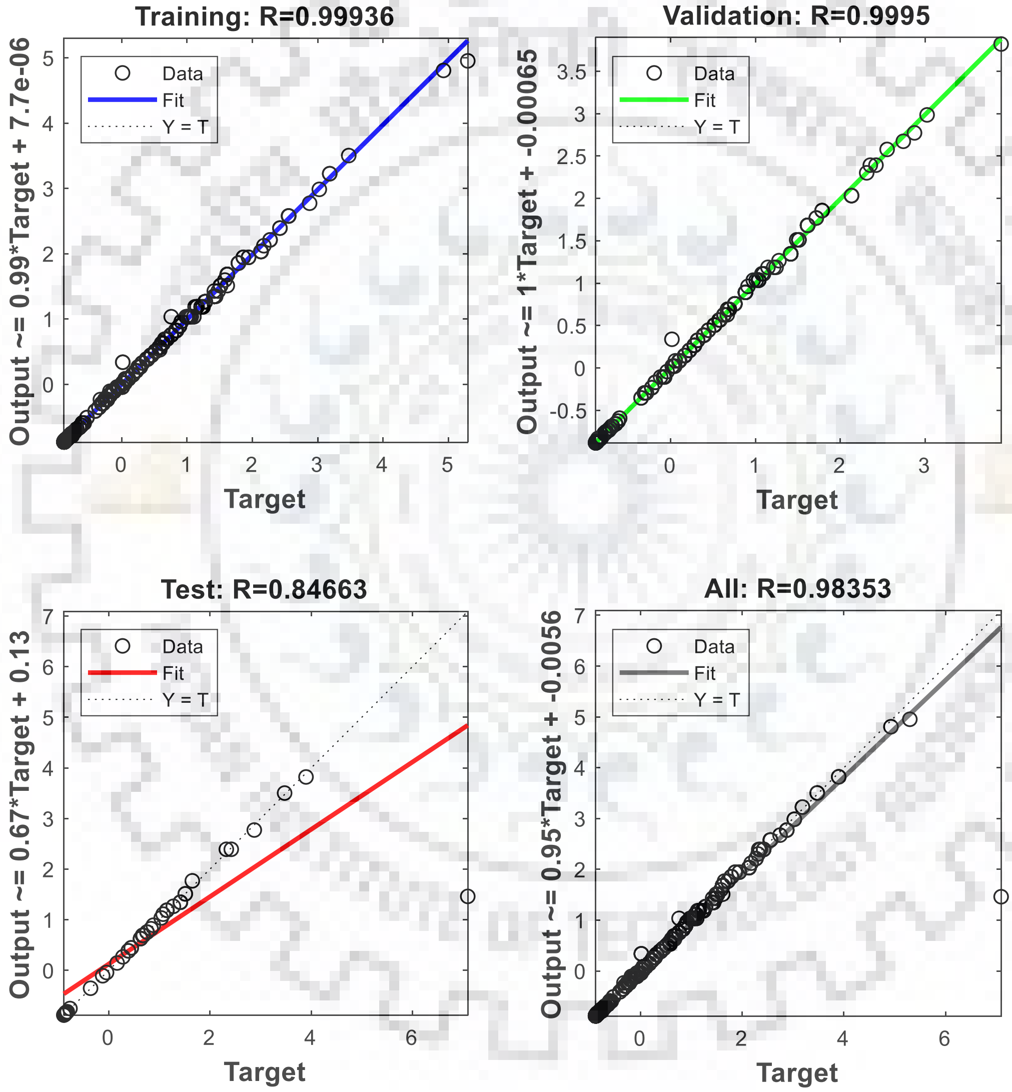


Figure 4.12. Representing Bhilangana river calibration and validation of Model 1.

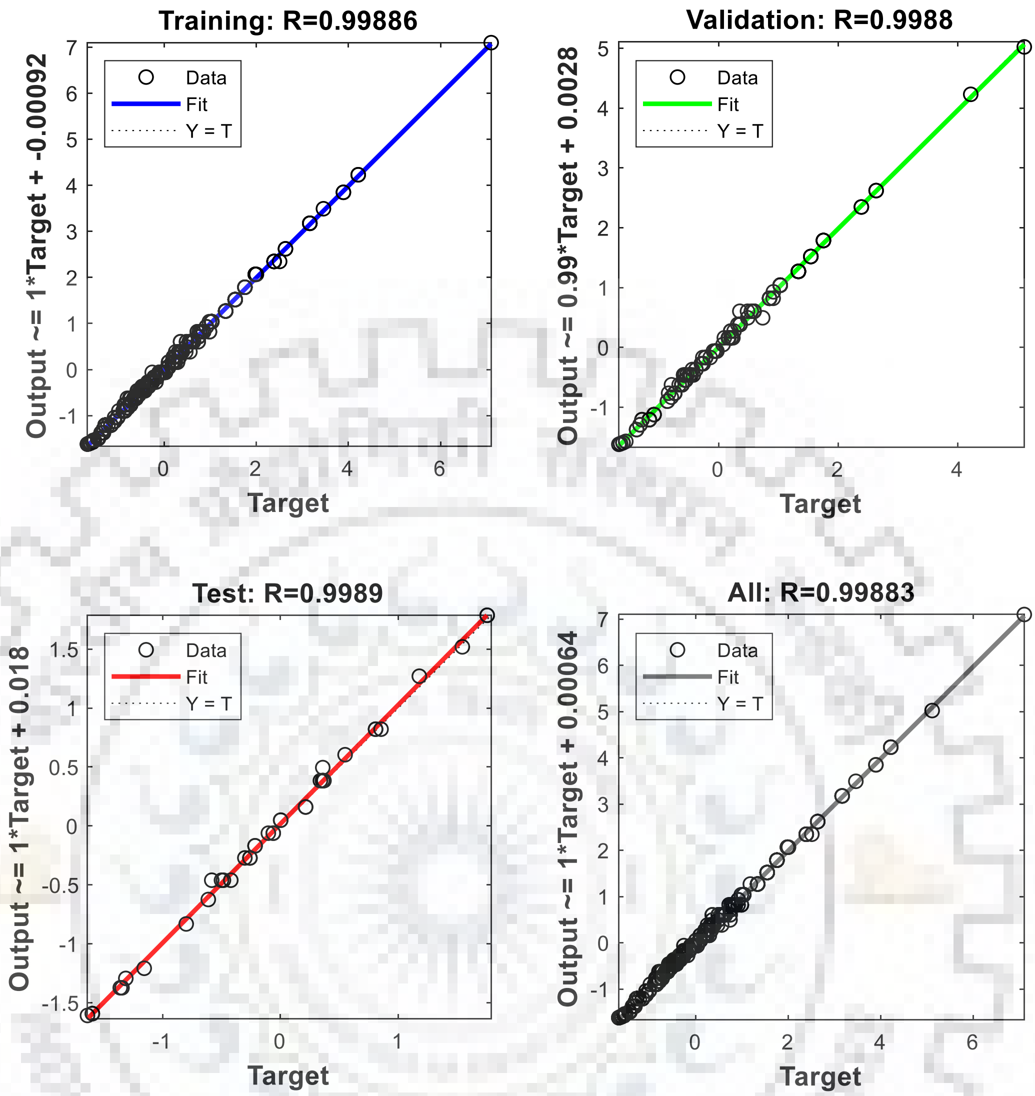


Figure 4.13. Representing Balganga river calibration and validation of Model 1.

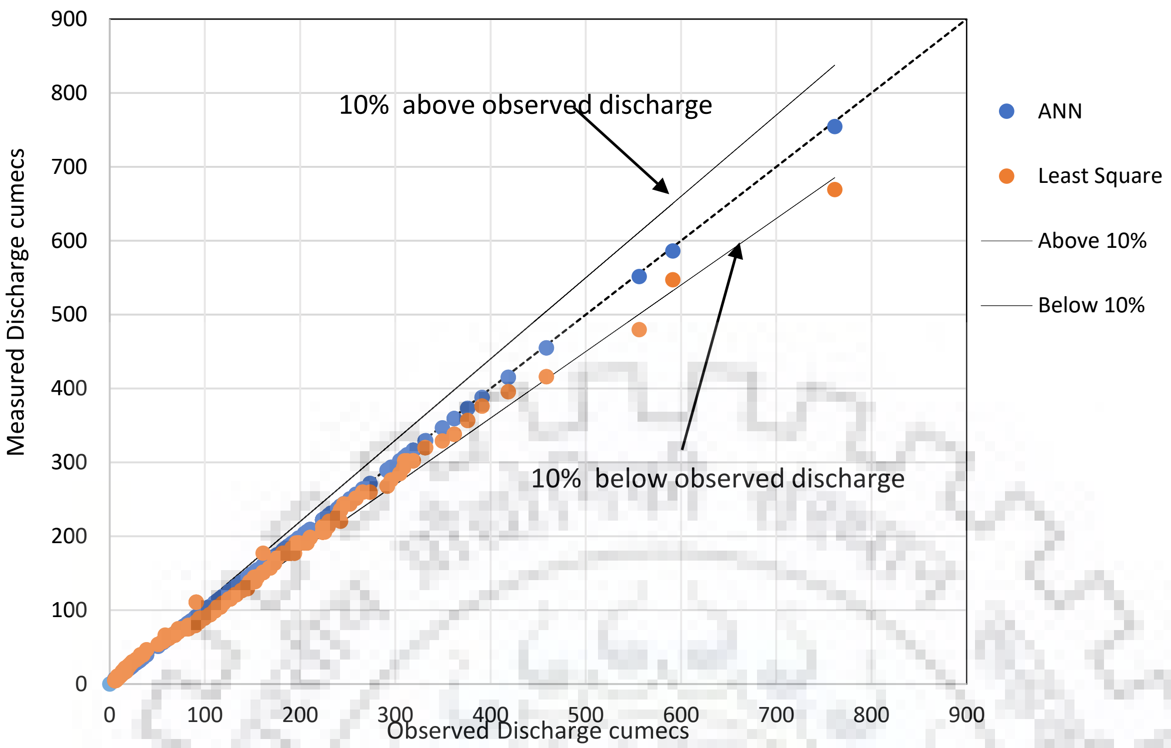


Figure 4.14.Observed and computed discharge a scatter plot for Bhilangana River at Ghansali without outliers

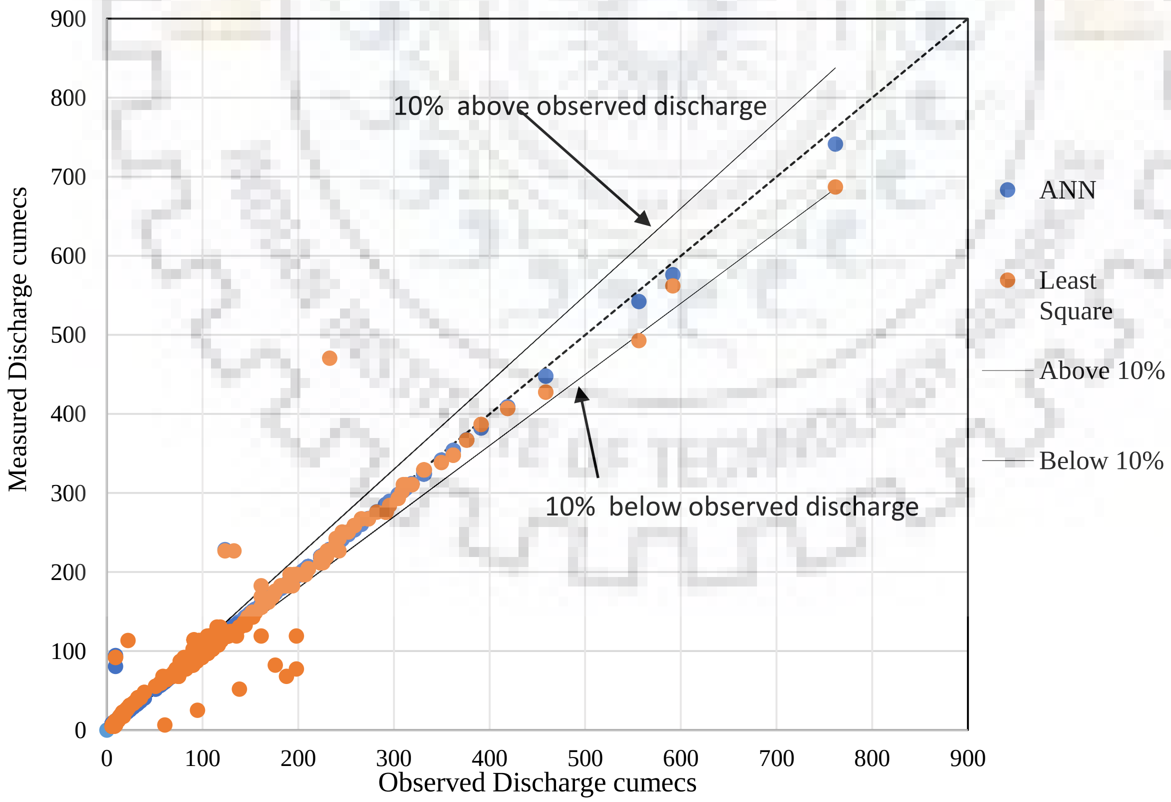


Figure 4.15. Observed and computed discharge a scatter plot for Bhilangana River at Ghansali with outliers

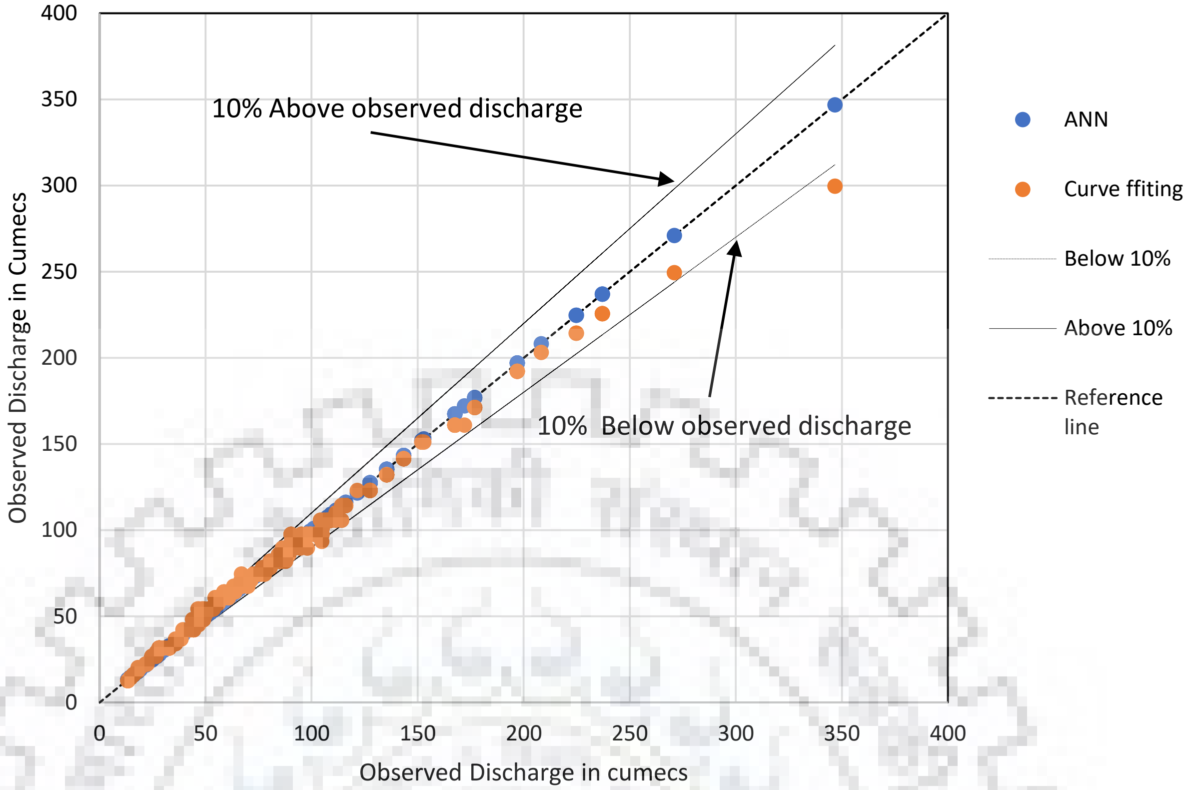


Figure 4.16. Observed and computed discharge a scatter plot for Balganga River at Sarasgaon without outliers

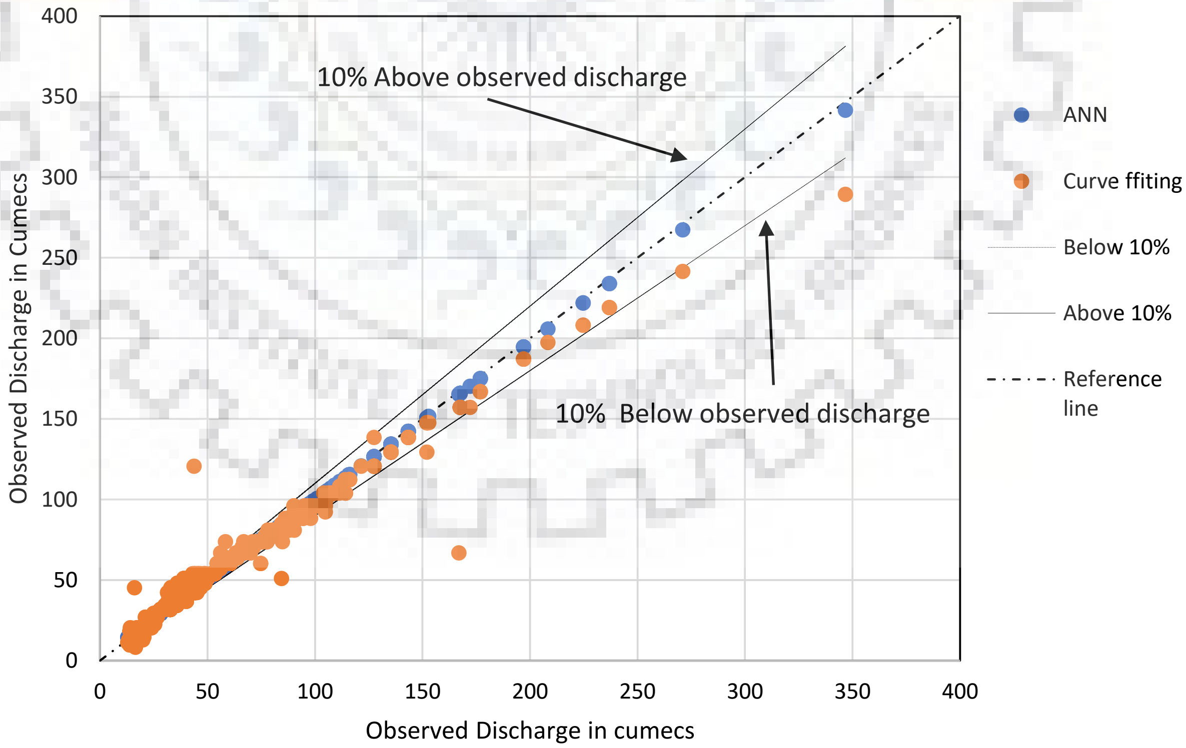


Figure 4.17. Observed and computed discharge a scatter plot for Balganga River at Sarasgaon without outliers

CHAPTER 5

DEVELOPMENT OF STOCHASTIC MODELS FOR DAILY STREAMFLOW FORECASTING

5.1 GENERAL

Hydrologists often deal with the limited number of observed data while analysing the time series (TS). Use of stochastic models can be a possible solution for that case as it does not consider the physical nature of the time series during modelling (Box and Jenkins, 1976; Shahin et al.,1993). In hydrological field, the stochastic models commonly used are: pure random (or white noise) model, autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model, autoregressive moving average model with exogenous inputs (ARMAX) and autoregressive integrated moving average (ARIMA) model. In the present study, the AR, ARX, ARMA and ARMAX models have been used for daily streamflow forecasting purpose.

This chapter presents details of data used, development of stochastic models for daily streamflow forecasting for monsoon and non-monsoon seasons, and their results. All the mathematical calculations of the stochastic model development have been done in Microsoft Excel software version 2010 and code was written in R-programming language using R studio version 3.4.3. The developed code of the stochastic model is attached in Appendix I.

5.2 STUDY AREA

Three sub-basins, namely Bhagirathi (at MBII), Bhilangana (at Ghansali), and Balganga (at Sarasgaon) of Tehri catchment are considered for this study. The details of all the sub-basins have already been described in Chapter 3.

5.2.1 Data Used

The rainfall and discharge data from three sub-basins, namely Bhagirathi at MB II, Bhilangana at Ghansali and Balganga at Sarasgaon were collected from Real-time inflow forecasting system website of Tehri dam. The availability of rainfall and discharge data is given in Table 5.1. Table 5.2 to Table 5.4 shows the statistical summary of the data set used for the present study for all three sub-basins.

Table 5.1.Observed discharge and stage data summary

Catchment	Test	Seasons		Data of Discharge, Rainfall
		Monsoon	Non-Monsoon	
Bhilangana	Calibration	June 2016 – September 2016 June 2017 to September 2017	October 2016 – May 2017; October 2017-May 2018	All
	Validation	June 2018 – September 2018	October 2018– May 15, 2019	All
Balganga	Calibration	June 2016 – September 2016 June 2017 to September 2017	October 2016 – May 2017; October 2017-May 2018	All
	Validation	June 2018 – September 2018	October 2018– May 15, 2019	All
Bhagirathi	Calibration	June 2016 – September 2016 June 2017 to September 2017	October 2016 – May 2017; October 2017-May 2018	All
	Validation	June 2018 – September 2018	October 2018– May 15, 2019	All

The equations used for computing the basic statistical characteristic of time series, like mean, sample variance, skewness coefficient, and standardization of the series data set are given below:

Sample Mean

$$\overline{Q} = \frac{1}{N}\sum_{t=1}^{N} Q_t \tag{5.1}$$

Where, N representing the length of sample size, and $\overline{Q}$ is mean sample of data set.

Sample Variance

$$S^2 = \frac{1}{N-1}\sum_{t=1}^{N}(Q_t - \overline{Q})^2 \tag{5.2}$$

Where, N representing the length of sample size, S is a variance of data set, Q_t is t-th data set series and $\overline{Q}$ is mean sample of data set. $t = 1, 2, 3, \ldots\ldots$

Skewness coefficient

$$\boldsymbol{\delta} = \frac{\frac{\mathbf{1}}{\boldsymbol{N}-\mathbf{1}} \sum_{t=1}^{N} (\boldsymbol{Q_t} - \overline{\boldsymbol{Q}})^2}{\boldsymbol{S}^3} \tag{5.3}$$

Where, N representing the length of sample size, δ is skewness coefficient, S is a variance of the data set, Q_t is t-th data set series and $\overline{Q}$ is mean sample of data set. $t = 1, 2, 3, \ldots\ldots$

Standardized series

$$\boldsymbol{Z_{v,\tau}} = \frac{\boldsymbol{Q_{v,\tau}} - \overline{\boldsymbol{Q}}_{\boldsymbol{\tau}}}{\boldsymbol{\sigma_\tau}} \tag{5.4}$$

Where, $Z_{v,\tau}$ is the standardized data set, σ_τ is the standard deviation of the data set, $Q_{v,\tau}$ length of the data set, and $\overline{Q}_\tau$ a sample mean of data set.

Table 5.2. Bhagirathi River monsoon data set analysis summary

No.	Parameter	The monsoon season from June to September			Non-monsoon season from October to May (next year)		
		2016	**2017**	**2018**	**2016**	**2017**	**2018**
1	Mean daily Runoff (Cumecs)	338.94	286.46	285.61	46.84	46.32	71.31
2	Standard Deviation	155.05	149.89	121.15	31.56	20.47	22.37
3	Coefficient of skewness	0.32	0.44	0.69	1.50	1.29	0.05
4	Coefficient of Kurtosis	-0.72	-1.24	0.15	1.36	0.60	-0.79
5	Max daily Runoff	715.49	592.80	687.20	162.28	110.70	117.00
6	Min daily Runoff	106.97	104.73	114.70	11.06	22.30	32.00

Table 5.3.Bhilangana River monsoon and non-monsoon data set analysis summary

No.	Parameter	The monsoon season from June to September			Non-monsoon season from October to May (next year)		
		2016	**2017**	**2018**	**2016**	**2017**	**2018**
1	Mean daily Runoff (Cumecs)	105.21	325.57	111.55	12.89	14.55	11.91
2	Standard Deviation	87.38	150.25	80.85	8.10	11.87	6.18
3	Coefficient of skewness	2.17	0.41	1.31	1.39	1.87	1.71
4	Coefficient of Kurtosis	7.63	-0.17	2.50	1.57	3.22	2.00
5	Max daily Runoff	564.75	786.85	446.00	43.13	60.46	30.82
6	Min daily Runoff	21.29	89.11	21.29	4.67	5.37	7.55

Table 5.4. Balganga River monsoon and non-monsoon data set analysis summary

No.	Parameter	The monsoon season from June to September			Non-monsoon season from October to May (next year)		
		2016	2017	2018	2016	2017	2018
1	Mean daily Runoff (Cumecs)	51.48	68.17	54.56	3.48	2.57	3.81
2	Standard Deviation	45.79	51.52	48.28	1.61	1.72	2.64
3	Coefficient of skewness	1.40	0.67	0.79	1.32	3.17	2.10
4	Coefficient of Kurtosis	2.73	0.61	0.27	1.53	9.18	4.84
5	Max daily Runoff	257.41	240.90	222.46	9.55	10.00	15.00
6	Min daily Runoff	3.54	5.00	2.00	2.00	2.00	0.00

5.3 STOCHASTIC MODELLING FORMULATION

AR and ARMA models are one of the important and popular stochastic models used for the time series analysis and forecasting. In the present study, AR, ARX, ARMA and ARMAX models with the exogenous variable inputs were developed for monsoon season. However, for non-monsoon season, only AR model is used to simulate the daily streamflow. The following are the mathematical expression of the models:

AR (p) model is represented as

$$Q_t = \varphi_1 Q_{t-1} + \ldots\ldots + \varphi_p Q_{t-p} + \varepsilon_t \tag{5.5}$$

ARMA (p, q) model;

$$Q_t = \varphi_1 Q_{t-1} + \ldots\ldots + \varphi_p Q_{t-p} + \Theta_1 \varepsilon_{t-1} + \ldots\ldots + \Theta_q \varepsilon_{t-q} + \ldots\ldots + \varepsilon_t \tag{5.6}$$

ARX (p, r) model;

$$Q_t = \varphi_1 Q_{t-1} + \ldots\ldots + \varphi_p Q_{t-p} + b_1 d_{t-1} + \ldots\ldots + b_r d_{t-r} + \varepsilon_t \tag{5.7}$$

ARMAX (p, q, r) model;

$$Q_t = \varphi_1 Q_{t-1} + \ldots\ldots + \varphi_p Q_{t-p} + \Theta_1 \varepsilon_{t-1} + \ldots\ldots + \Theta_q \varepsilon_{t-q} + b_1 d_{t-1} + \ldots\ldots + b_r d_{t-r} + \varepsilon_t \tag{5.8}$$

Where, Qt is the time dependent series (Variable), φ_1 to φ_p are the coefficients of AR terms, Θ_1 to Θ_q are the coefficients of MA, b_1 to b_q are the coefficients of exogenous input variable (Rainfall or Temperature).

5.3.1 Time-series modelling procedure

The development of stochastic model involves three steps procedures, namely the model identification, the model parameter estimation and the model diagnostic checking. In this study, after the development procedure, the appropriate model was selected, which can produce good results for simulating daily flow by using the historical streamflow pattern, rainfall and temperature data.

5.3.1.1 Model identification

The identification of a model is the initial stage in order to exemplify the behaviour of the time series (TS) and to estimate the order of the model (p and q). In the present study, the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) have been used for the identification of the model order. The *Eqs. 6.9 to 6.12* have been used to compute ACF and PACF of the residuals. Table 5.5 shows the conditions for the identification of model order for TS models.

C_k is usually called the lag-k autocovariance,

$$C_k = \frac{1}{N}\sum_{t=1}^{N-k}(Q_t - \overline{Q})(Q_{t+k} - \overline{Q}), \qquad 0 \le K < N \tag{5.9}$$

Where, K represents lag time (or distance) between correlated pairs (Q_t, Q_{t+k}), N is total number of sample size, $\overline{Q}$ is the average sample.

For a certain case that K = 0, C_0 turns into the variance S^2 of the Eq. (6.2)

$$C_0 = \frac{1}{N}\sum_{t=1}^{N}(Q_t - \overline{Q})^2, \qquad 0 \le K < N \tag{5.10}$$

$$r_k = \frac{C_k}{C_0} = \frac{\sum_{t=1}^{N-k}(Q_t - \overline{Q})(Q_{t+k} - \overline{Q})}{\sum_{t=1}^{N}(Q_t - \overline{Q})^2} \tag{5.11}$$

$$\mathrm{r_k}(95\%) = \frac{-1 \pm 1.96\sqrt{\mathrm{N - K - 1}}}{\mathrm{N - K}} \tag{5.12}$$

Where, r_k is named the lag-k autocorrelation coefficient, the lag-k is the serial correlation coefficient of the autocorrelation function (ACF). The plot of r_k against k is termed as the correlogram.

Table 5.5.Identification of the ACF and PACF for AR, MA and ARMA

Models	ACF	PACF
AR (p)	Decays geometrically	P significant lags (order)
MA (q)	P significant lags (order)	Decays geometrically
ARMA (p, q)	Decays geometrically	Decays geometrically

5.3.1.2 Parameter estimation

The method of moments and the method of maximum likelihood is the two statistical methods usually used to estimate the parameters (Box and Jenkins, 1976; Salas et al., 1980). In the present study, the method of moments has been used to compute the parameters of the model. The expressions of the method of moments are as follows:

a) For autoregressive with exogenous variable input ARX (p, r) model;

$$Q_t = \varphi_1 Q_{t-1} + \ldots\ldots + \varphi_p Q_{t-p} + b_1 d_{t-1} + \ldots\ldots + b_r d_{t-r} + \varepsilon_t \tag{5.13}$$

$$Q_{t+1} = \varphi_1 Q_{t+1-1} + \ldots\ldots + \varphi_p Q_{t+1-p} + b_1 d_{t+1-1} + \ldots\ldots + b_r d_{t+1-r} + \varepsilon_t \tag{5.14}$$

$$Q_{t+T} = \varphi_1 Q_{t+T-1} + \ldots\ldots + \varphi_p Q_{t+T-p} + b_1 d_{t+T-1} + \ldots\ldots + b_r d_{t+T-r} + \varepsilon_t \tag{5.15}$$

Where, φ_p is the p-th autoregressive coefficient of the AR(p) model, b_r is the r-th exogenous variable coefficient of the X(r).

AR(p), p = 1, 2, 3……… X(r), r = 1, 2, 3………

Which may be written as matrix notation

$$\begin{bmatrix} Q_t \\ \vdots \\ \vdots \\ Q_{t+T} \end{bmatrix} = \begin{bmatrix} Q_{t-1} & d_{t-1} \\ \vdots & \vdots \\ \vdots & \vdots \\ Q_{t+T-1} & d_{t+T-1} \end{bmatrix} \begin{bmatrix} \varphi_p \\ \vdots \\ \vdots \\ b_r \end{bmatrix} \tag{5.16}$$

Where,

$$Y = \begin{bmatrix} Q_t \\ \vdots \\ \vdots \\ Q_{t+T} \end{bmatrix}, \; H = \begin{bmatrix} Q_{t-1} & d_{t-1} \\ \vdots & \vdots \\ \vdots & \vdots \\ Q_{t+T-1} & d_{t+T-1} \end{bmatrix}, \; \Theta = \begin{bmatrix} \varphi_p \\ \vdots \\ \vdots \\ b_r \end{bmatrix} \tag{5.17}$$

$$[Y] = [H]\,[\Theta]$$

$$[H^T]\,[Y] = [H^T]\,[H]\,[\Theta] \tag{5.18}$$

Then, parameter [Θ] is

$$[\Theta] = [H^T \times H]^{-1}[H^T][Y] \quad (5.19)$$

Thus the parameter [Θ] array is determined by the applying equation 6.19.

b) The linear equation for ARMAX with exogenous variable input (p, q, r) model are;

$Q_t = \varphi_1 Q_{t-1} + \ldots\ldots + \varphi_p Q_{t-p} + \Theta_1 \varepsilon_{t-1} + \ldots\ldots + \Theta_q \varepsilon_{t-q} + b_1 d_{t-1} + \ldots\ldots + b_r d_{t-r} + \varepsilon_t$

$Q_{t+1} = \varphi_1 Q_{t+1-1} + \ldots\ldots + \varphi_p Q_{t+1-p} + \Theta_1 \varepsilon_{t+1-1} + \ldots\ldots + \Theta_q \varepsilon_{t+1-q} + b_1 d_{t+1-1} + \ldots\ldots + b_r d_{t+1-r} + \varepsilon_t$

$Q_{t+T} = \varphi_1 Q_{t+T-1} + \ldots\ldots + \varphi_p Q_{t+T-p} + \Theta_1 \varepsilon_{t+T-1} + \ldots\ldots + \Theta_q \varepsilon_{t+T-q} + b_1 d_{t+T-1} + \ldots\ldots + b_r d_{t+T-r} + \varepsilon_t$

Where, φ_p is the p-th autoregressive coefficient of the AR(p) model, Θ_q is the q-th moving average coefficient of the MA(q), b_r is the r-th exogenous variable coefficient of the X(r).

AR(p), p = 1, 2, 3………. MA(q), q = 1, 2, 3………. X(r), r = 1, 2, 3……….

Which may be written as matrix notation

$$\begin{bmatrix} Q_t \\ \vdots \\ \vdots \\ \vdots \\ Q_{t+T} \end{bmatrix} = \begin{bmatrix} Q_{t-1} & \varepsilon_{t-1} & d_{t-1} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ Q_{t+T-1} & \varepsilon_{t+T-1} & d_{t+T-1} \end{bmatrix} \begin{bmatrix} \varphi_p \\ \vdots \\ \Theta_q \\ \vdots \\ b_r \end{bmatrix} \quad (5.20)$$

Where,

$$Y = \begin{bmatrix} Q_t \\ \vdots \\ \vdots \\ \vdots \\ Q_{t+T} \end{bmatrix}, \; H = \begin{bmatrix} Q_{t-1} & \varepsilon_{t-1} & d_{t-1} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ Q_{t+T-1} & \varepsilon_{t+T-1} & d_{t+T-1} \end{bmatrix}, \; \Theta = \begin{bmatrix} a_p \\ \vdots \\ \Theta_q \\ \vdots \\ b_r \end{bmatrix} \quad (5.21)$$

$$[Y] = [H][\Theta]$$

$$[H^T][Y] = [H^T][H][\Theta] \quad (5.22)$$

Then, parameter [Θ] is

$$[\Theta] = [H^T \times H]^{-1}[H^T][Y] \quad (5.23)$$

Thus the parameter [Θ] array is determined by the applying equation 6.23.

5.3.1.3 Model diagnostic checking

In the present study, the model diagnostic checking was done by using three-step procedures. At first, the model was tested by means of goodness of fit using the Autocorrelation in model residuals, i.e. by ACF and PACF. Thereafter, the model performance was checked by using different model performance indicators (NSE, RMSE, R^2 and MSE), and finally, the best model was chosen on the basis of lowest Akaike Information Criteria (AIC), provided other indices were also either the best ones or were close to the best ones.

AIC value of a model can be estimated by using the following formula:

AIC for an AR(p)

$$AIC\ (p) = N(LN(SE^2)) + 2p \tag{5.24}$$

Where, SE^2 is the residual variance of standardized series, N total number of samples and p is the parameters of models.

And, AIC for an ARMA (p, q)

$$AIC\ (p, q) = N\ (LN(SE^2)) + 2(p + q) \tag{5.25}$$

Where, SE^2 is the residual variance of standardized series, N total number of samples, and (p, q) is the parameters of models.

5.3.2 Model performance

The performance of a model can be assessed by using different performance indicators. In the present study, three different performance indicators were used, i.e. correlation coefficient (r), the root means square error (RMSE) and Nash Sutcliffe efficiency (NSE) (Nash & Sutcliffe, 1970). The evaluation of the models are given by the following equations:

$$\text{The coefficient of Correlation} = \frac{\sum_{i=1}^{N}(Q_O-\bar{Q}_O)(Q_p-\bar{Q}_p)}{\sqrt{\sum_{i=1}^{N}(Q_O-\bar{Q}_O)^2\sum_{i=1}^{N}(Q_p-\bar{Q}_p)^2}} \tag{5.26}$$

$$\text{Root Mean Square Error} = \sqrt{\frac{\sum_{i=1}^{N}(Q_{Oi}-\bar{Q}_{pi})^2}{N}} \tag{5.27}$$

$$\text{Nash Sutcliffe Efficiency} = 100 \times \left[1 - \frac{\sum_{i=1}^{N}(Q_O-\bar{Q}_p)^2}{\sum_{i=1}^{N}(Q_O-\bar{Q}_O)^2}\right] \tag{5.28}$$

where, N is the Number of observations; Q_o is the observed flow, Q_p is the predicted flow, $\overline{Q}_O$ is the mean of the observed flow and $\overline{Q}_p$ is the mean of the predicted flow.

5.4 RESULTS AND DISCUSSION

This section discusses results obtained after using the stochastic models for the three sub-catchments of the Tehri basin.

5.4.1 ACF and PACF plot

The model is identified using the ACF and PACF plot in the present study. The ACF and PACF plots for all three sub-catchments are shown in Figure 5.1 (a to c). In this figure, the first plot represents to ACF, and the second one represents to PACF plot. From Figure 5.1 (a and c), the ACF graph shows the decays geometrically and PACF values show significant in lag 1 and rest is non-significant. From Figure 5.1(b), ACF values show the decays geometrically, and PACF values are shows the significant in lag 1 and lag 2 and the rest in non-significant. According to the ACF and PACF results, AR and ARX models are appropriate for the time series.

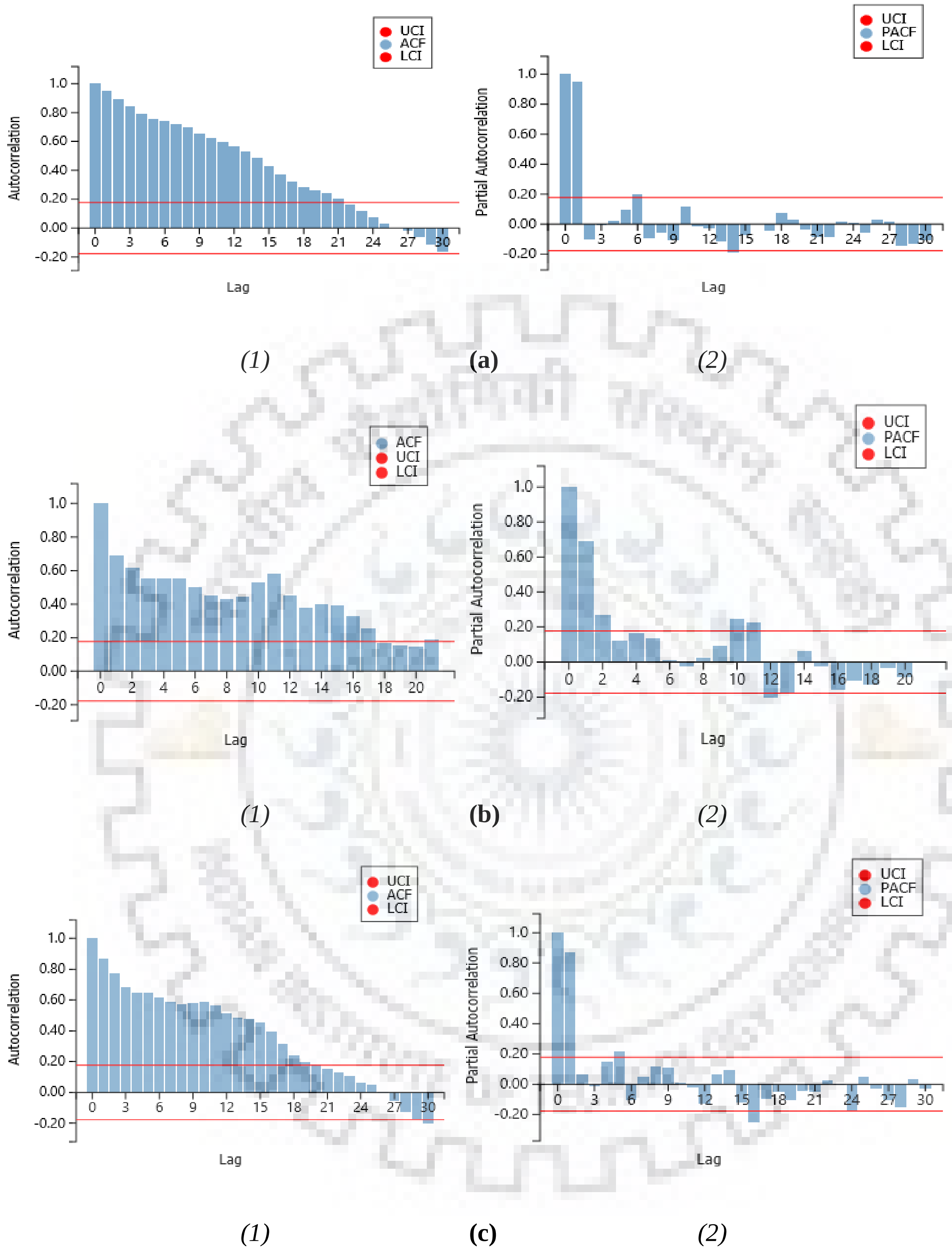


Figure 5.1. (a) Bhagirathi River (b) Bhilangana River and (c) Balganga River representing ACF and PACF values. *(red line representing confidence limits of the model ± 95%)*

5.4.2 Calibration performance of the Model

In the present study, the model was developed on a seasonal basis, i.e. for monsoon and non-monsoon season differently.

5.4.2.1 Monsoon model performance

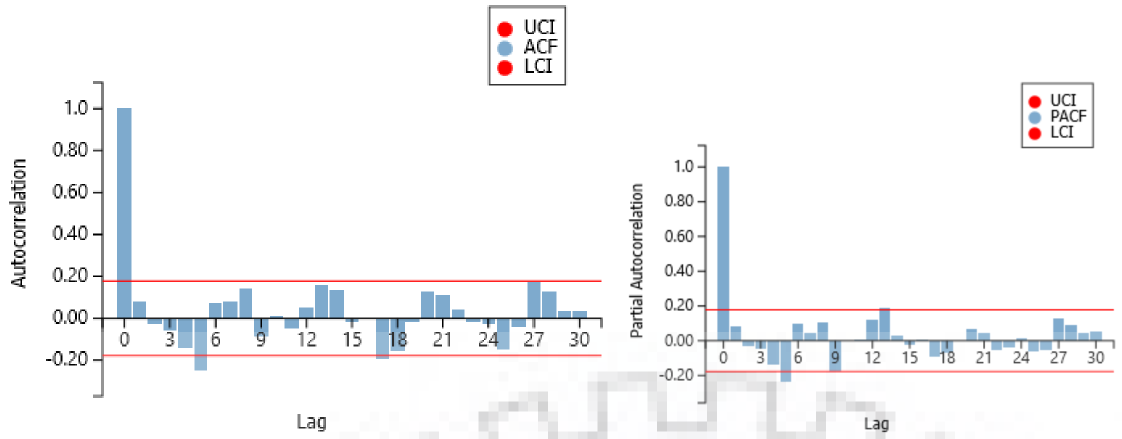
For monsoon season, four stochastic linear TS models, namely AR, ARX, ARMA and ARMAX models with exogenous variable inputs have been developed for this study. The model calibration results are given in Table 5.6 to Table 5.8 for the three sub-basin, namely Bhagirathi, Bhilangana and Balganga, respectively. The value of Nash and Sutcliffe Efficiency (NSE), Coefficient of Determination (R2), Root Mean Square Error (RMSE), and Akaike Information Criteria (AIC) are given in the tables.

The results of the Bhagirathi sub-basin are shown in Table 5.6. On the basis of the lowest AIC value, the results showed that the ARX (1, 0, 1) model is better than the other models. The lowest value of AIC is -283.517. The model performance indicator of the selected model shows very good results with high values of NSE (0.970), PBIAS% (0.537) and coefficient of determination (0.985). The graph of ACF and PACF of residuals is shown in Figure 5.2 for Bhagirathi sub-basin. The results show that the ACF and PACF residual values are falling within the confidence limit, which indicated the acceptance of the selected TS model. The observed discharge and the model simulated discharge is shown in Figure 5.3 and also plotted for the Bhagirathi sub-basin. The results of observed and simulated model values are shows a clear match.

Table 5.6.Calibration monsoon Bhagirathi Catchment June 2016 to September 2016

No.	Models	Parameters	NSE	RMSE	MAE	PBIAS %	R2	AIC
1	AR	(1,0,0)	0.898	49.435	34.542	0.824	0.956	-269.740
2	AR	(2,0,0)	0.897	49.561	34.614	0.830	0.956	-266.720
3	AR	(3,0,0)	0.904	47.912	33.733	0.676	0.959	-266.867
4	ARMA	(1,1,0)	0.977	23.166	17.514	-0.309	0.990	-217.870
5	ARMA	(2,1,0)	0.974	24.680	16.898	-0.310	0.989	-91.552
6	ARMA	(3,1,0)	0.976	24.228	17.054	-0.283	0.990	-40.404
7	**ARX**	**(1,0,1)**	**0.970**	**27.945**	**20.436**	**0.537**	**0.985**	**-283.517**
8	ARMAX	(1,1,1)	0.990	15.792	11.952	0.375	0.995	-245.156
9	ARMAX	(2,1,1)	0.970	26.623	19.044	-0.532	0.986	-143.033
10	ARMAX	(3,1,1)	0.972	26.032	18.804	-0.543	0.987	-121.580
12	ARMAX	(1,1,2)	0.973	25.583	20.040	-0.683	0.987	-222.211
13	ARMAX	(2,1,2)	0.970	26.606	18.943	-0.503	0.986	-142.036

*The bold row indicates the chosen model for calibration.



(1) *(2)*

Figure 5.2. ACF and PACF residual of the model ARX(1,0,1) Bhagirathi river (2014). *(red line representing confidence limits of the model ± 95%)*

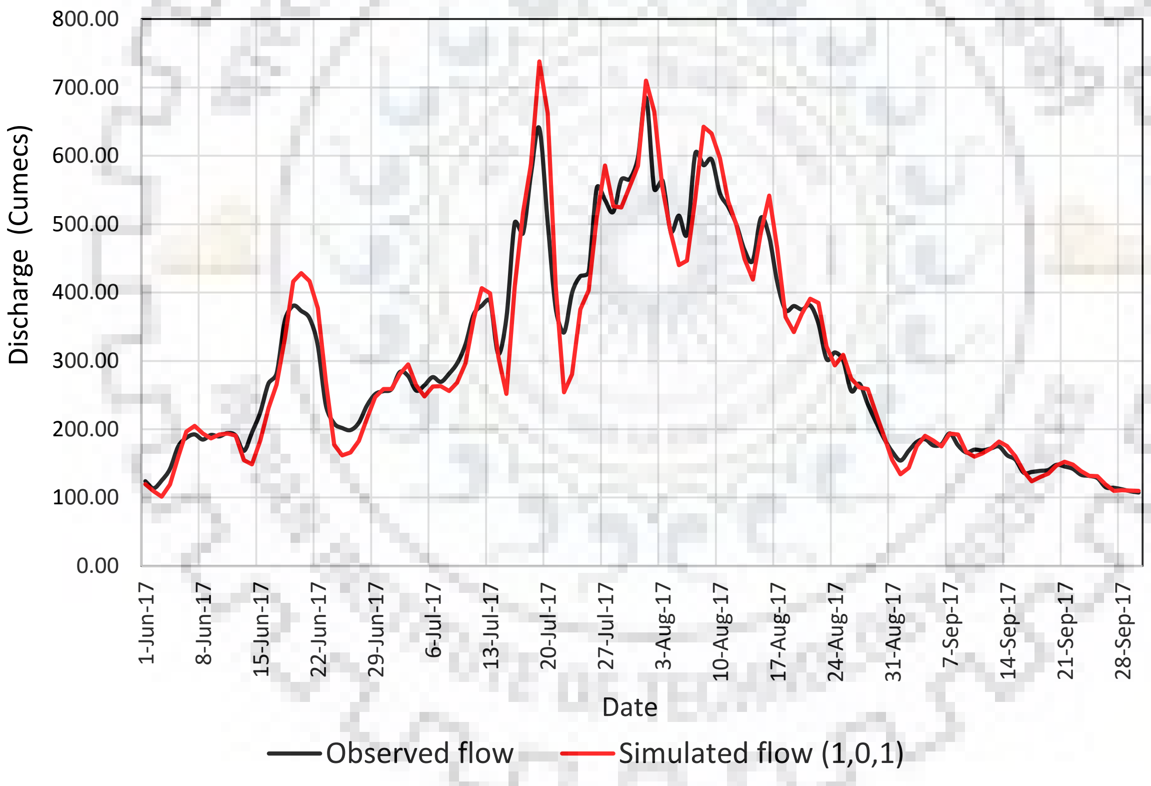


Figure 5.3 Calibration monsoon Bhagirathi Catchment June 2017 to September 2017

The results of the Bhilangana sub-basin are shown in Table 5.7. On the basis of the lowest AIC value, the results showed that the ARMAX (1, 1, 1) model is better than the other models. The lowest value of AIC is -154.510. The model performance indicator of the selected model shows very good results with high values of NSE (0.910), PBIAS% (0.827) and coefficient of determination (0.958). The graph of ACF and PACF of residuals is shown in Figure 5.4 for Bhilangana sub-basin. The results show that the ACF and PACF residual values are falling within the confidence limit, which indicated the acceptance of the selected TS model. The observed discharge and the model simulated discharge is shown in Figure 5.5 and also plotted for the Bhilangana sub-basin. The results of observed and simulated model values are shows a clear match.

Table 5.7.Calibration monsoon Bhilangana Catchment June 2016 to September 2016

No.	Models	Parameters	NSE	RMSE	MAE	PBIAS %	R^2	AIC
1	AR	(1,0,0)	0.648	50.794	29.848	14.329	0.869	-61.358
2	AR	(2,0,0)	0.769	41.144	25.499	8.673	0.907	-72.733
3	AR	(3,0,0)	0.820	36.326	23.322	6.817	0.925	-73.845
4	ARMA	(1,1,0)	0.980	12.215	6.768	-0.116	0.991	-137.373
5	ARMA	(2,1,0)	0.980	12.216	6.884	-0.114	0.991	-134.742
6	ARMA	(3,1,0)	0.977	13.278	7.265	-0.134	0.989	-62.201
7	ARX	(1,0,1)	0.776	48.095	28.738	12.983	0.893	-34.051
8	**ARMAX**	**(1,1,1)**	**0.910**	**31.038**	**21.523**	**0.827**	**0.958**	**-154.510**
9	ARMAX	(2,1,1)	0.977	13.053	7.172	-0.282	0.989	-130.010
10	ARMAX	(3,1,1)	0.975	13.656	7.397	-0.285	0.988	-105.367
12	ARMAX	(1,1,2)	0.857	32.349	22.031	-0.933	0.955	-149.528
13	ARMAX	(2,1,2)	0.857	32.369	22.084	-0.934	0.955	-147.205

*The bold row indicates the chosen model for calibration.

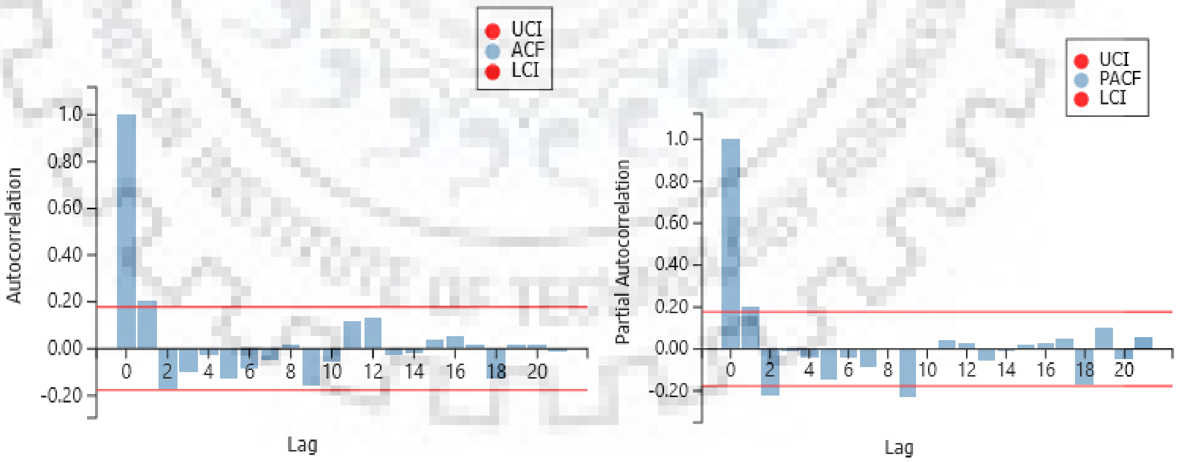


Figure 5.4.ACF and PACF residual of the model ARX(1,0,1) Bhilangana river (2017). *(red line representing confidence limits of the model ± 95%)*

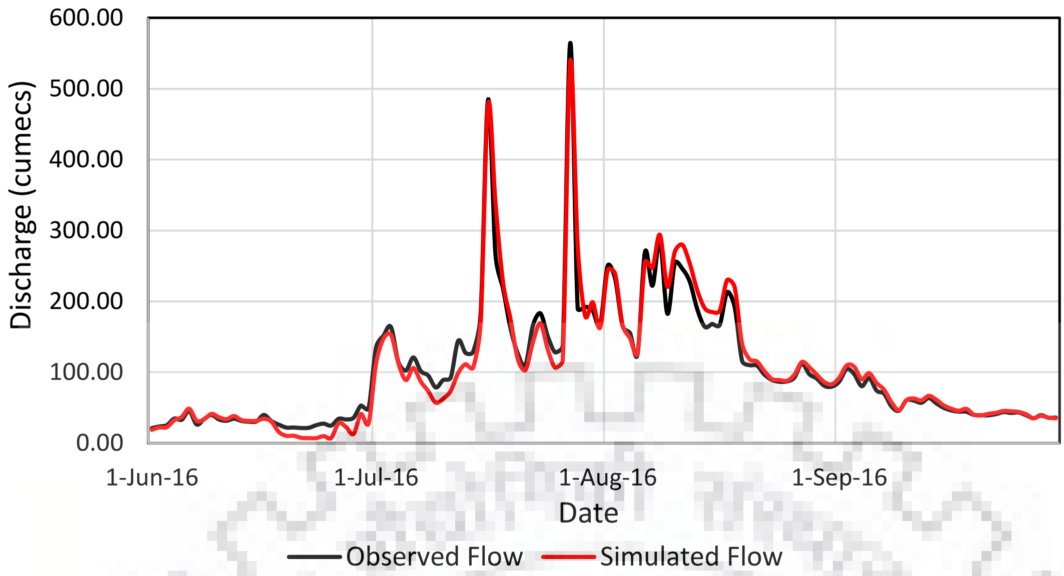


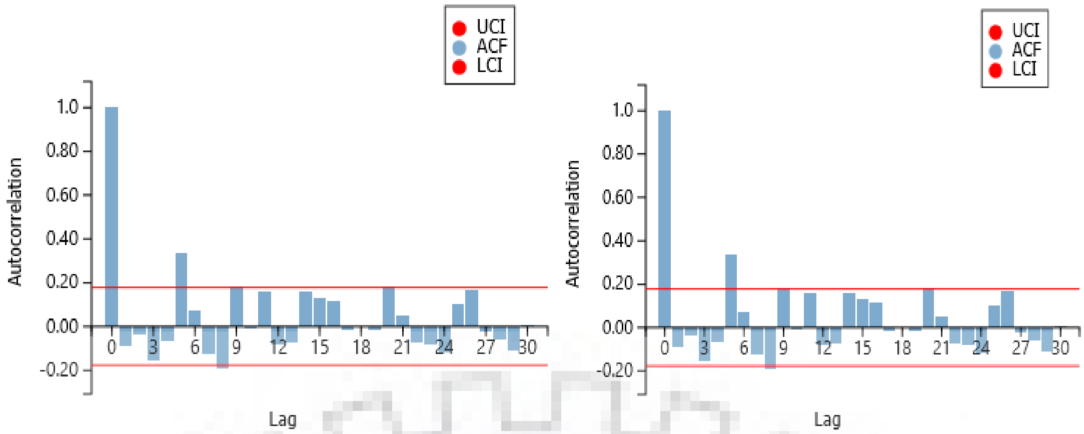
Figure 5.5. Calibration monsoon Bhilangana Catchment June 2016 to September 2016

The results of the Balaganga sub-basin are shown in Table 5.8. On the basis of the lowest AIC value, the results showed that the ARMAX (1, 1, 1) model is better than the other models. The lowest value of AIC is -188.701. The model performance indicator of the selected model shows very good results with high values of NSE (0.971), PBIAS% (0.319) and coefficient of determination (0.986). The graph of ACF and PACF of residuals is shown in Figure 5.6 for Balaganga sub-basin. The results show that the ACF and PACF residual values are falling within the confidence limit, which indicated the acceptance of the selected TS model. The observed discharge and the model simulated discharge is shown in Figure 5.7 and also plotted for the Balaganga sub-basin. The results of observed and simulated model values are shows a clear match.

Table 5.8.Calibration monsoon Balganga Catchment June 2016 to September 2016

No.	Models	Parameters	NSE	RMSE	MAE	PBIAS %	R^2	AIC
1	AR	(1,0,0)	0.704	24.793	16.612	6.054	0.906	-176.585
2	AR	(2,0,0)	0.720	24.143	16.422	5.428	0.910	-174.531
3	AR	(3,0,0)	0.723	24.016	16.280	5.291	0.910	-171.500
4	ARMA	(1,1,0)	0.892	14.962	10.102	-0.133	0.962	-126.596
5	ARMA	(2,1,0)	0.893	14.926	9.887	-0.089	0.963	-108.271
6	ARMA	(3,1,0)	0.894	15.112	10.115	0.051	0.964	-100.613
7	ARX	(1,0,1)	0.967	9.840	6.964	2.752	0.985	-113.058
8	**ARMAX**	**(1,1,1)**	**0.971**	**9.384**	**6.385**	**0.319**	**0.986**	**-188.701**
9	ARMAX	(2,1,1)	0.878	15.916	10.548	-0.884	0.959	-172.126
10	ARMAX	(3,1,1)	0.879	15.846	10.459	-0.792	0.959	-149.161
12	ARMAX	(1,1,2)	0.854	17.406	12.353	-0.407	0.954	-112.894
13	ARMAX	(2,1,2)	0.853	17.462	12.183	-0.242	0.955	-96.799

*The bold row indicates the chosen model for calibration.



(1) (2)

Figure 5.6. ACF and PACF residual of the model ARMAX(1,1,1) Balganga (2016). *(red line representing confidence limits of the model ± 95%)*

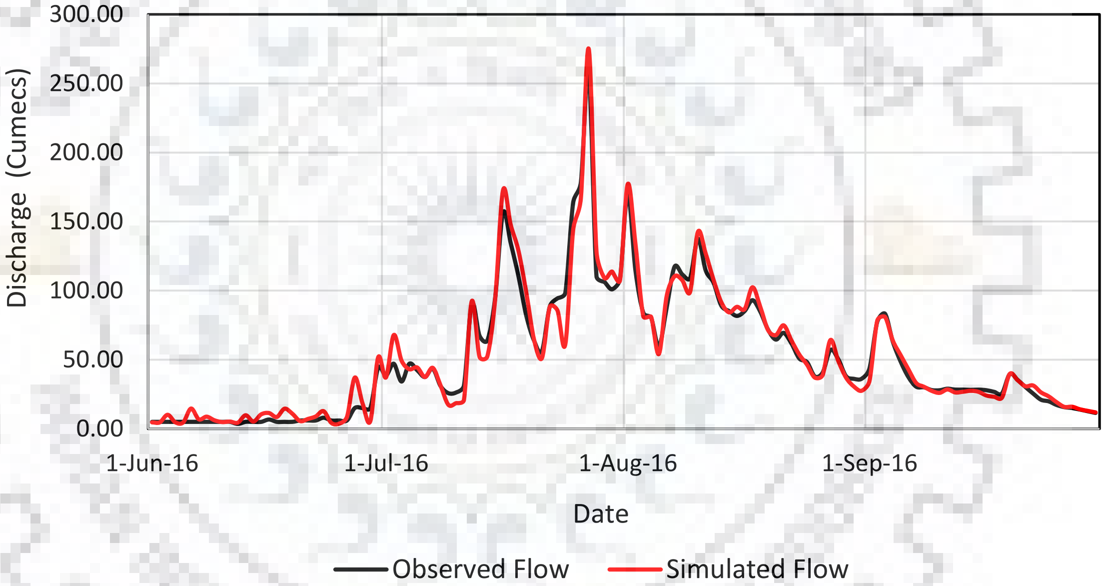


Figure 5.7. Calibration monsoon Balganga Catchment June 2016 to September 2016

5.4.2.2 Non-Monsoon model performance

The AR model is only used for the non-monsoon season. Therefore, the AR model results are discussed in this section. The calibration results for the non-monsoon season for all three sub-basins are given in Table 5.9. The table is describing the computed statistical analysis of the models such as Nash and Sutcliffe Efficiency (NSE), PBIAS%, Coefficient of Determination (R2), Root Mean Square Error (RMSE).

For the Bhagirathi sub-basin, the model performance indicator indicated a very high value of NSE. The results also indicate the suitability of the model on the basis of lower PBIAS% and higher value of coefficient of determination (Table 5.9). The observed discharge and the model simulated discharge is shown in Figure 5.8 and also plotted for the Bhagirathi sub-basin. The results of observed and simulated model values are shows a clear match. The graph of ACF and PACF of residuals is shown in Figure 5.9 for Bhagirathi sub-basin. The results show that the ACF and PACF residual values are falling within the confidence limit, which indicated the acceptance of the selected TS model.

Table 5.9. Calibration performance results in Non-monsoon

Catchments	Year	Models	NSE	RMSE	MAE	PBIAS %	R^2
Bhagirathi MB II	2016/2017	AR(1)	0.939	7.804	4.454	7.763	0.976
	2017/2018	AR(1)	0.904	6.510	4.345	8.153	0.965
Bhilangana	2016/2017	AR(1)	0.953	1.756	1.006	2.582	0.978
	2017/2018	AR(1)	0.980	1.665	0.846	-0.188	0.991
Balganga	2016/2017	AR(1)	0.967	0.292	0.156	1.738	0.985
	2017/2018	AR(1)	0.962	0.337	0.090	-1.096	0.986

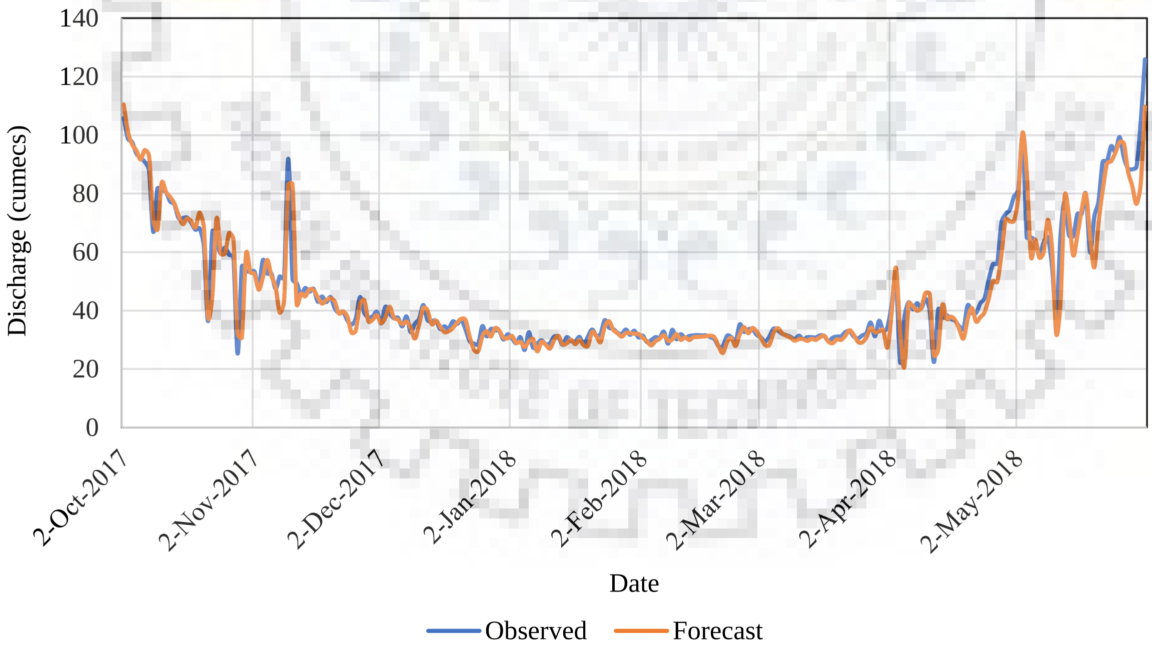
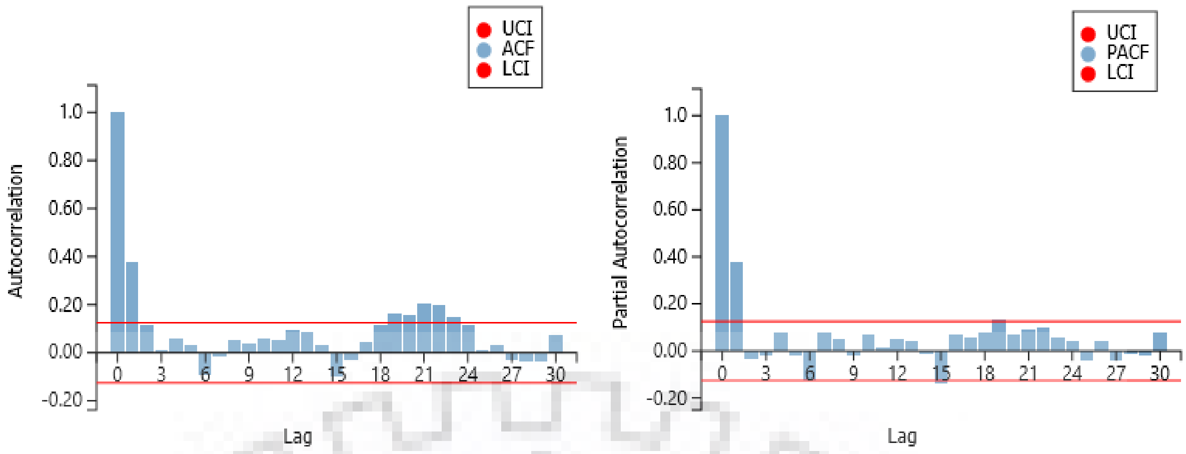


Figure 5.8. Calibration non-monsoon Bhagirathi Catchment October 2017 to May 2018



(1) *(2)*

Figure 5.9. Non-monsoon ACF and PACF residual of the model AR(1) Bhagirathi river (2017). *(red line representing confidence limits of the model ± 95%)*

For the Bhilangana sub-basin, the model performance indicator indicated a very high value of NSE. The results also indicate the suitability of the model on the basis of lower PBIAS% and higher value of coefficient of determination (Table 5.9). The observed discharge and the model simulated discharge is shown in Figure 5.10 and also plotted for the Bhilangana sub-basin. The results of observed and simulated model values are shows a clear match. The graph of ACF and PACF of residuals is shown in Figure 5.11 for Bhilangana sub-basin. The results show that the ACF and PACF residual values are falling within the confidence limit, which indicated the acceptance of the selected TS model.

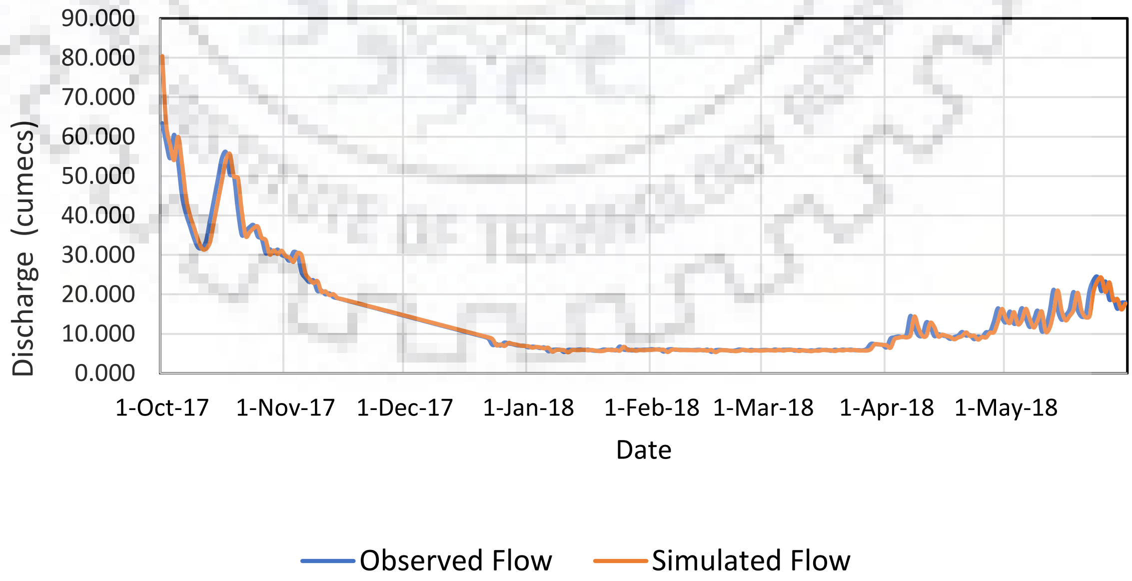
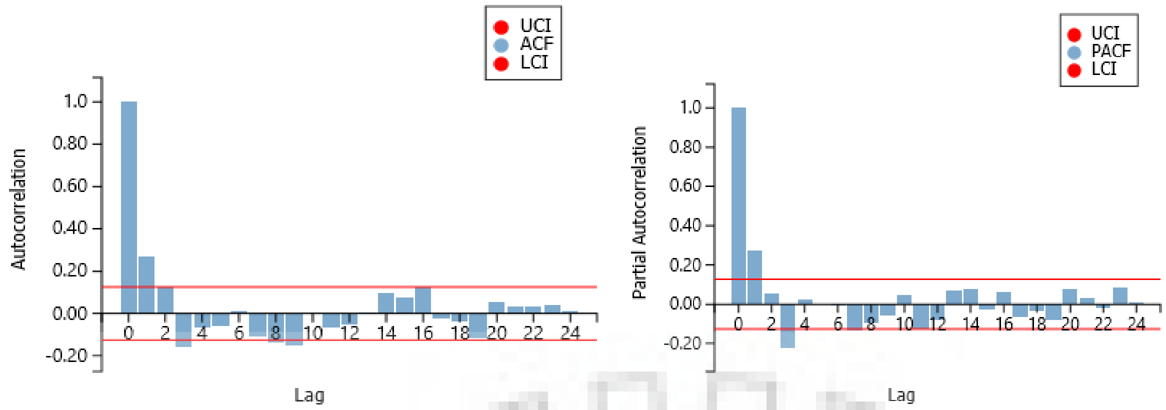


Figure 5.10. Calibration non-monsoon Bhilangana Catchment October 2017 to May 2018



(1) *(2)*

Figure 5.11. Non-monsoon ACF and PACF residual of the model AR(1) Bhilangana river (2016). *(red line representing confidence limits of the model ± 95%)*

For the Balganga sub-basin, the model performance indicator indicated a very high value of NSE (0.967). The results also indicate the suitability of the model on the basis of lower PBIAS% (1.738) and higher value of coefficient of determination (0.985) (Table 5.9). The observed discharge and the model simulated discharge is shown in Figure 5.12 and also plotted for the Balganga sub-basin and. The results of observed and model simulated values are shows a clear match. The graph of ACF and PACF of residuals is shown in Figure 5.13 for Balganga sub-basin. The results show that the ACF and PACF residual values are falling within the confidence limit, which indicated the acceptance of the selected TS model.

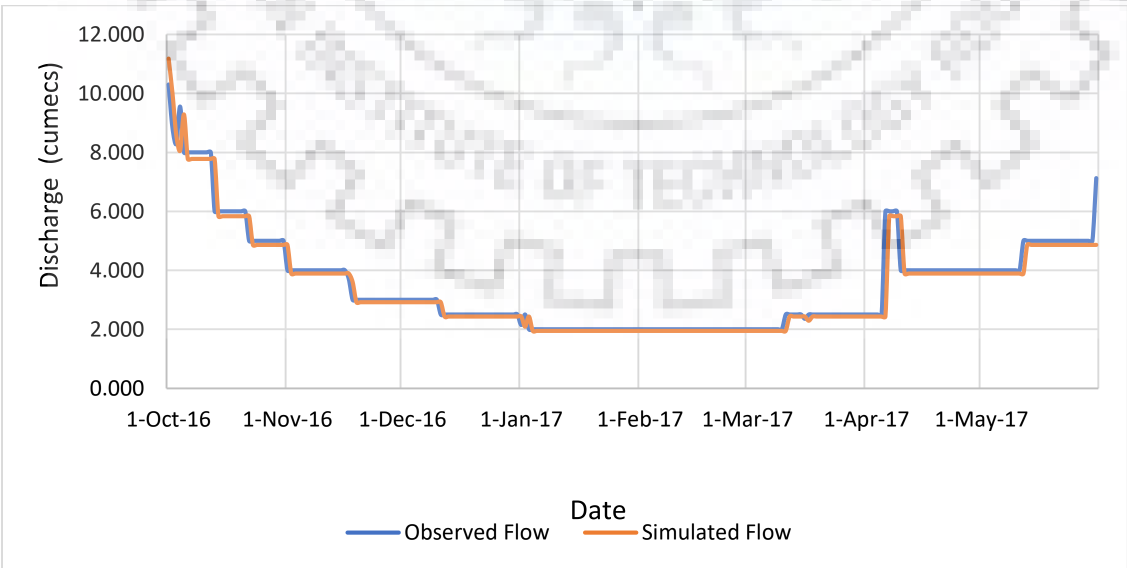
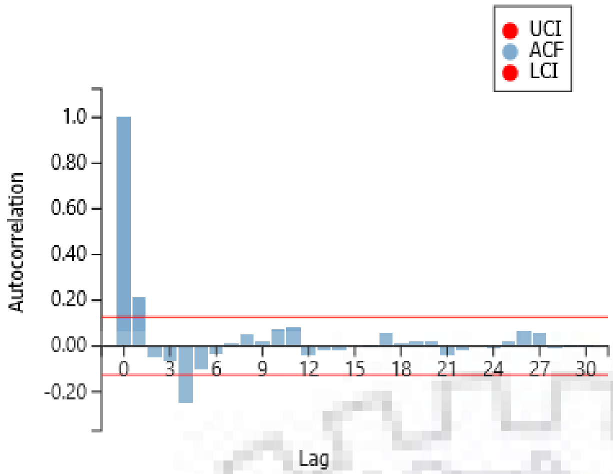
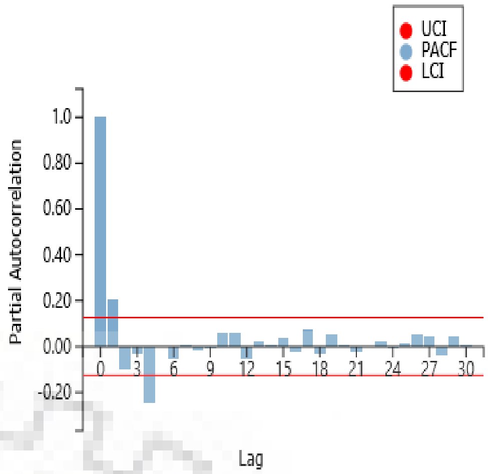


Figure 5.12. Calibration non-monsoon Balganga Catchment October 2016 to May 2017



(1)



(2)

Figure 5.13. Non-monsoon ACF and PACF residual of the model AR(1) Balganga (2016).

(red line representing confidence limits of the model ± 95%)

5.4.3 Validation performance of the Model

5.4.3.1 Monsoon model performance

The results of model validation in monsoon season for Bhagirathi, Bhilangana and Balganga sub-basins are shown in Table 5.10. For Bhagirathi sub-basin, the values of NSE, PBIAS% and coefficient of determination are 0.934, -2.147 and 0.976, respectively. The statistical analysis results show the satisfaction of model. The observed streamflow and model simulated streamflow are compared and shown in Figure 5.14. The results of observed and simulated streamflow are shows the similarity, which confirms that the selected model can be used for the forecasting purpose during the monsoon season for the Bhagirathi sub-basin.

Table 5.10. Validation performance results in monsoon from June 2018 to September 2018

Models	Models	NSE	RMSE	MAE	PBIAS %	Coefficient R^2
Bhagirathi	ARX (1,0,1)	0.934	34.482	25.648	-2.147	0.976
Bhilangana	ARMAX(1,1,1)	0.894	27.239	20.410	-2.491	0.946
Balganga	ARMAX(1,1,1)	0.966	10.348	6.481	0.013	0.983

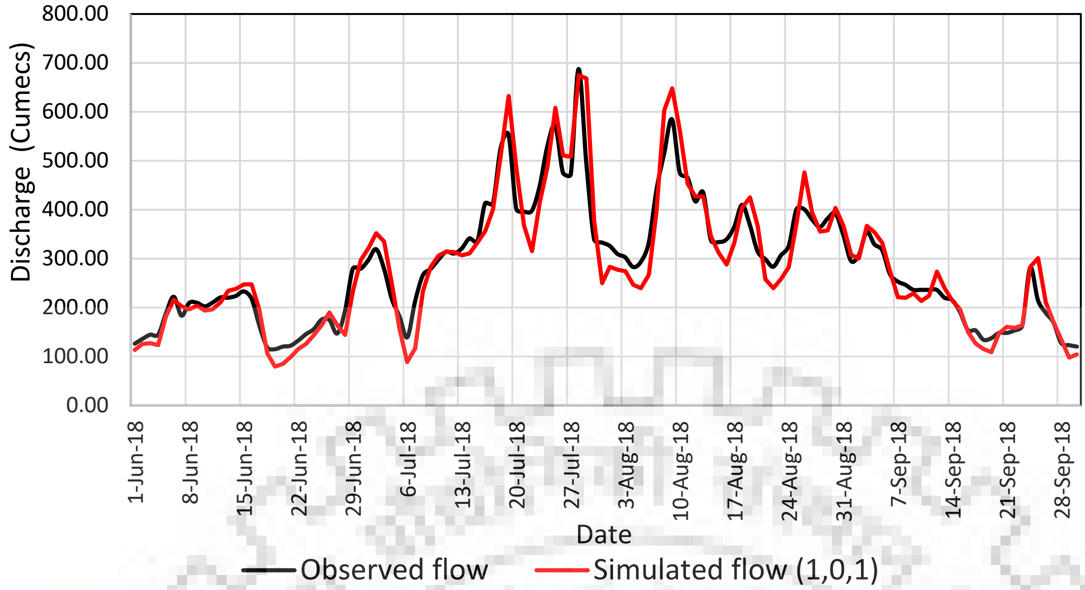


Figure 5.14. Validation monsoon Bhagirathi Catchment June 2018 to September 2018 using 2016 parameters

For Bhilangana sub-basin, the values of NSE, PBIAS% and coefficient of determination are 0.894, -2.491 and 0.946, respectively (Table 5.10). The statistical analysis results show the satisfaction of model. The observed streamflow and model simulated streamflow are compared and shown in Figure 5.15. The results show the similarity between observed and simulated streamflow, which confirms that the selected model can be used for the forecasting purpose during the monsoon season for the Bhilangana sub-basin.

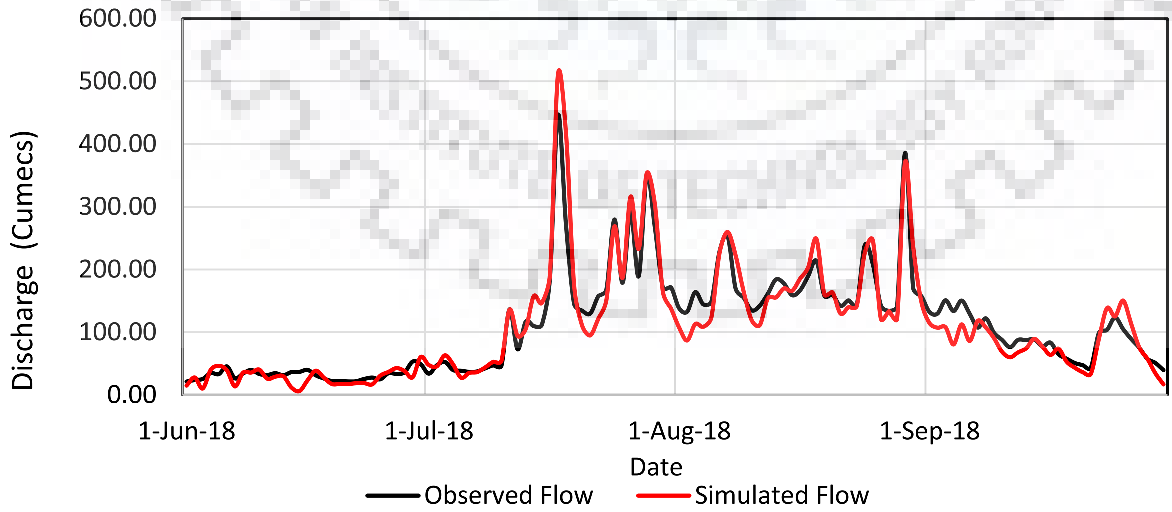


Figure 5.15. Validation monsoon Bhilangana Catchment June to Sept 2018 using 2017 parameters

For Balganga sub-basin, the values of NSE, PBIAS% and coefficient of determination are 0.966, 0.013 and 0.983, respectively (Table 5.10). The statistical analysis results show the satisfaction of model. The observed streamflow and model simulated streamflow are compared and shown in Figure 5.16. The results show the similarity between observed and simulated streamflow, which confirms that the selected model can be used for the forecasting purpose during the monsoon season for the Balganga sub-basin.

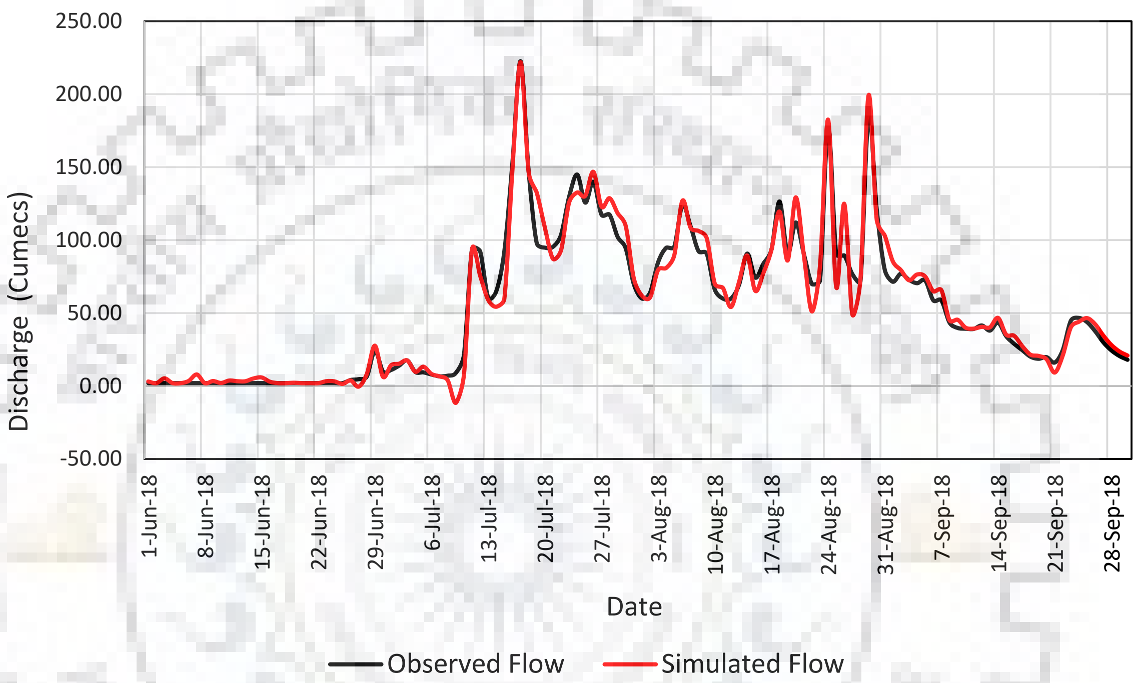


Figure 5.16. Validation monsoon Balganga Catchment June 2018 to September 2018 using 2016 parameters

5.4.3.2 Non-Monsoon model performance

The results of model validation in monsoon season for Bhagirathi, Bhilangana and Balganga sub-basins are shown in Table 5.11. For Bhagirathi sub-basin, the values of NSE, PBIAS% and coefficient of determination are 0.906, 6.850 and 0.965, respectively. The statistical analysis results show the satisfaction of model. The observed streamflow and model simulated streamflow are compared and shown in Figure 5.17. The results show the similarity between observed and simulated streamflow.

Table 5.11. Validation performance results in non-monsoon from October 2018 to May 2019

Catchments	Year	Models	NSE	RMSE	MAE	PBIAS %	Coefficient R^2
Bhagirathi	2018/2019	AR(1)	0.906	7.598	5.161	6.850	0.965
Bhilangana	2018/2019	AR(1)	0.965	1.365	0.810	1.143	0.983
Balganga	2018/2019	AR(1)	0.978	0.324	0.185	-0.309	0.991

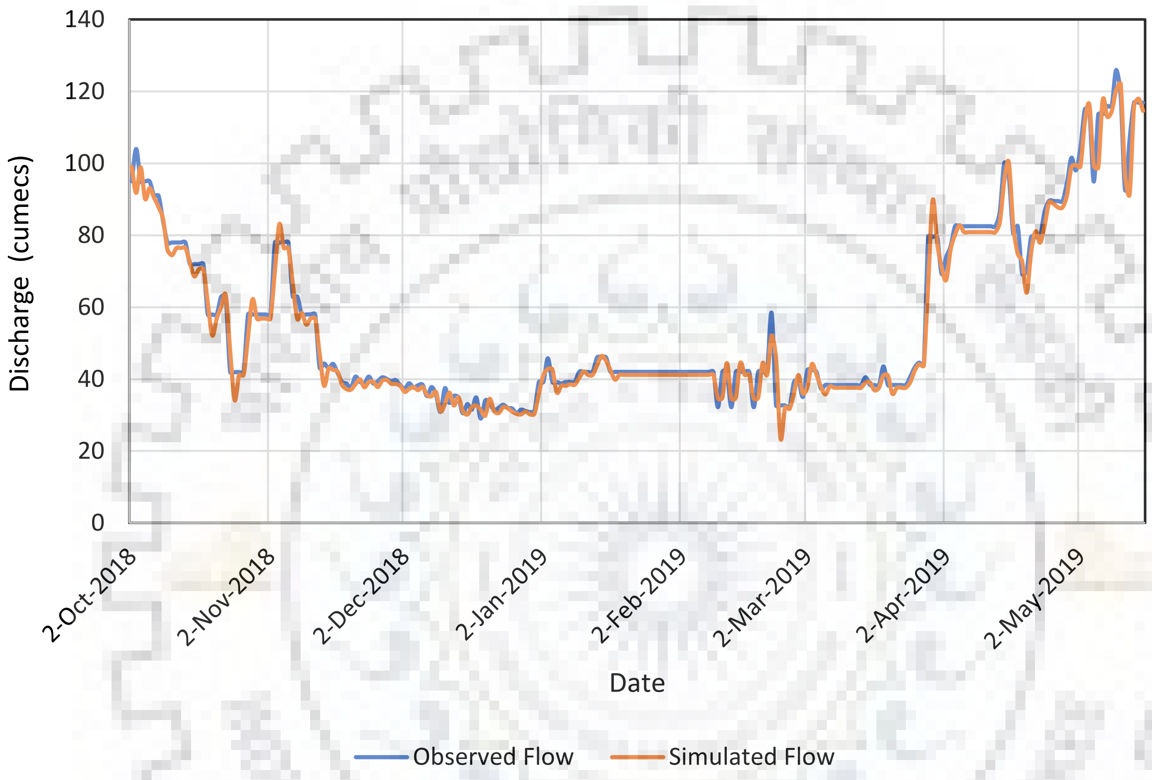


Figure 5.17. Validation Non-monsoon Bhagirathi Catchment October 2018 to May 2019

For Bhilangana sub-basin, the values of NSE, PBIAS% and coefficient of determination are 0.965, 1.143 and 0.983, respectively (Table 5.11). The statistical analysis results show the satisfaction of model. The observed streamflow and model simulated streamflow are compared and shown in Figure 5.18. The results show the similarity between observed and simulated streamflow.

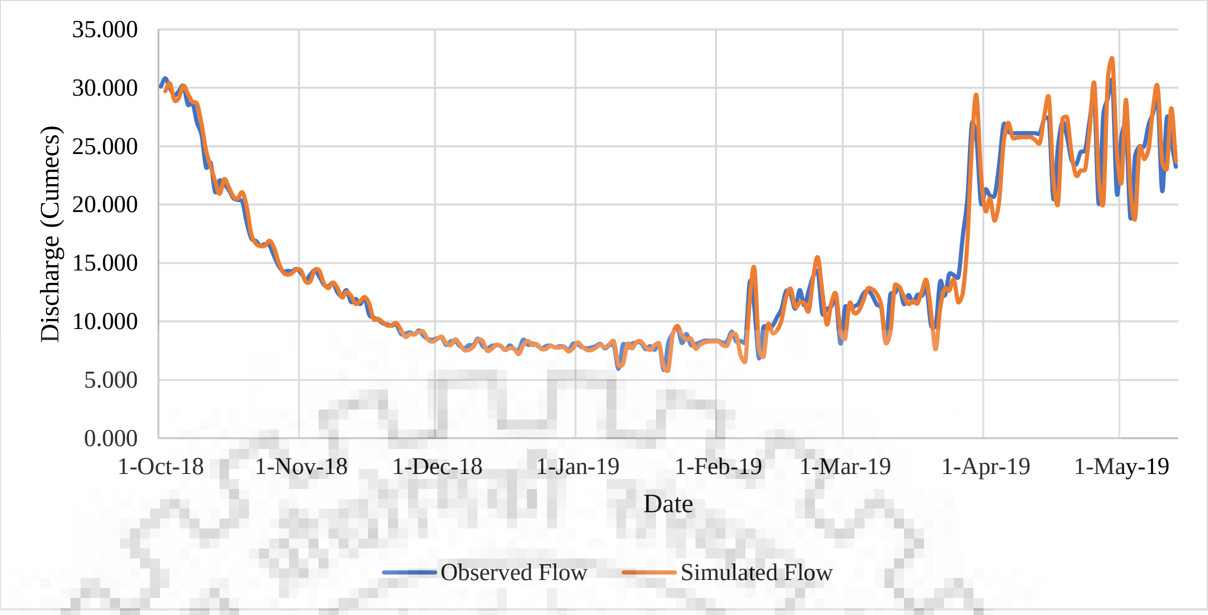


Figure 5.18. Validation Non-monsoon Bhilangana Catchment October 2018 to May 2019

For Balganga sub-basin, the values of NSE, PBIAS% and coefficient of determination are 0.978, -0.309 and 0.991, respectively (Table 5.11). The statistical analysis results show the satisfaction of model. The observed streamflow and model simulated streamflow are compared and shown in Figure 5.19. The results show the similarity between observed and simulated streamflow.

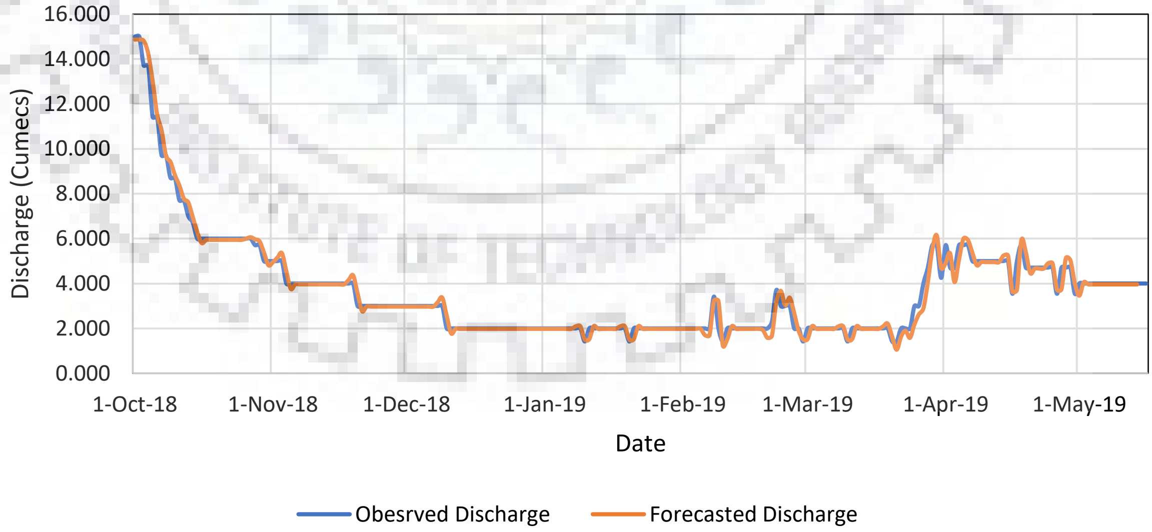


Figure 5.19. Validation Non-monsoon Balganga Catchment October 2018 to May 2019

5.5 COMPARISON OF STOCHASTIC AND HEC-HMS MODEL (BY AGRAWAL 2018)

The HEC-HMS model has been setup and used for the Bhagirathi, Bhilangana and Balganga sub-basins in an earlier study by Agrawal (2018). In the present study also the same three sub-basins were used for simulating daily streamflow. Therefore, the results of stochastic model and the HEC-HMS model were compared in the present study. The comparison results of Stochastic model and HEC-HMS model for Bhagirathi, Bhilangana and Balganga sub-basins in terms of model performance indicator are shown in Table 5.12 and Table 5.13 during calibration and validation process, respectively. For all three sub-basins, the results indicated that the performance (on the basis of model performance criteria and visual inspection) of stochastic models for calibration and validation are far better than the HEC-HMS model. The observed streamflow and simulated streamflow using stochastic models and HEC-HMS model are also compared and shown in Figure 5.20 to Figure 5.22 for Bhagirathi, Bhilangana and Balganga sub-basins. The results of Agrawal (2018) for HEC-HMS model were further cross verified by setting up the model again and making additional efforts to improve the model efficiency. The results of HEC-HMS model obtained in this study are presented in next section.

Table 5.12. Calibration performance results from June 2016 to May 2018

S/No.	Sub-Basins	NSE		Coefficient R^2		RMSE		PBIAS %	
		Stochastic Models	HEC-HMS	Stochastic Models	HEC-HMS	Stochastic Models	HEC-HMS	Stochastic Models	HEC-HMS
1	Bhagirathi MB II	0.982	0.752	0.991	0.885	20.754	77.124	-0.080	-4.204
2	Bhilangana	0.957	0.679	0.980	0.842	12.384	33.692	-1.114	-7.260
3	Balganga	0.924	0.587	0.965	0.776	10.755	25.127	0.263	20.724

Table 5.13. Validation performance results from June 2018 to Nov 2018

S/No.	Sub-Basins	NSE		Coefficient R^2		RMSE		PBIAS %	
		Stochastic Models	HEC-HMS	Stochastic Models	HEC-HMS	Stochastic Models	HEC-HMS	Stochastic Models	HEC-HMS
1	Bhagirathi MB II	0.984	0.758	0.992	0.870	18.289	97.447	-0.271	-6.185
2	Bhilangana	0.976	0.712	0.990	0.853	11.831	46.454	-0.607	-2.729
3	Balganga	0.840	0.717	0.902	0.793	20.726	26.739	-9.911	5.116

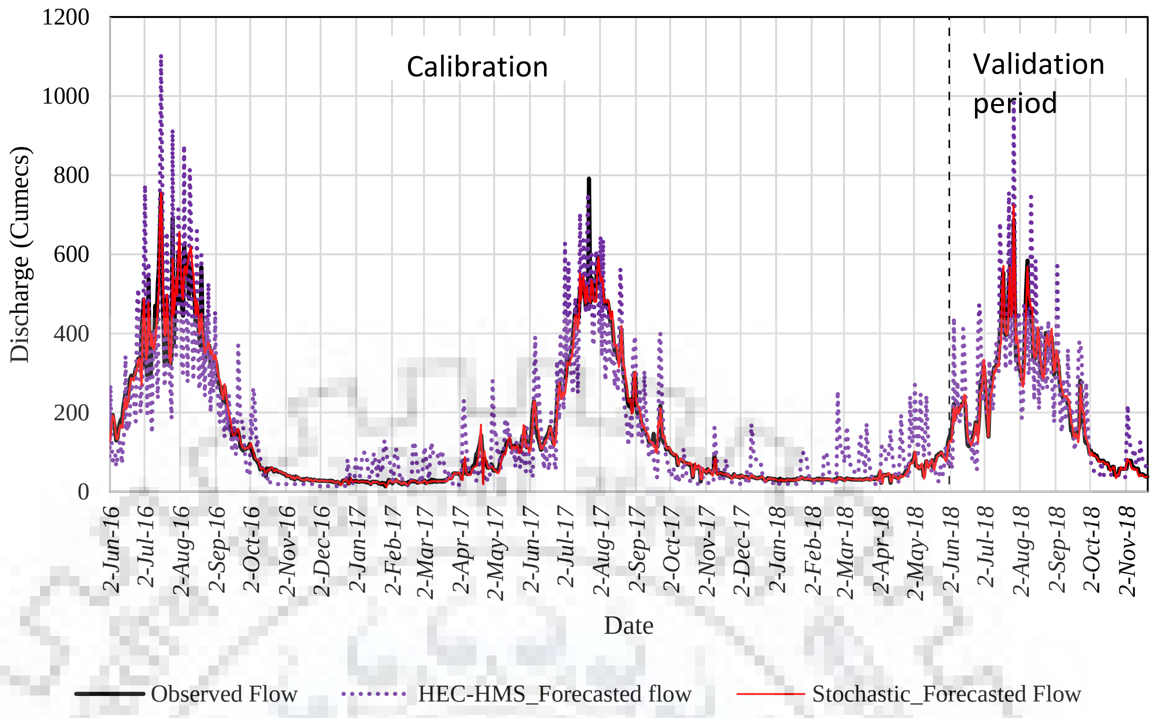


Figure 5.20. Observed and forecasted streamflow for Bhagirathi sub-basins

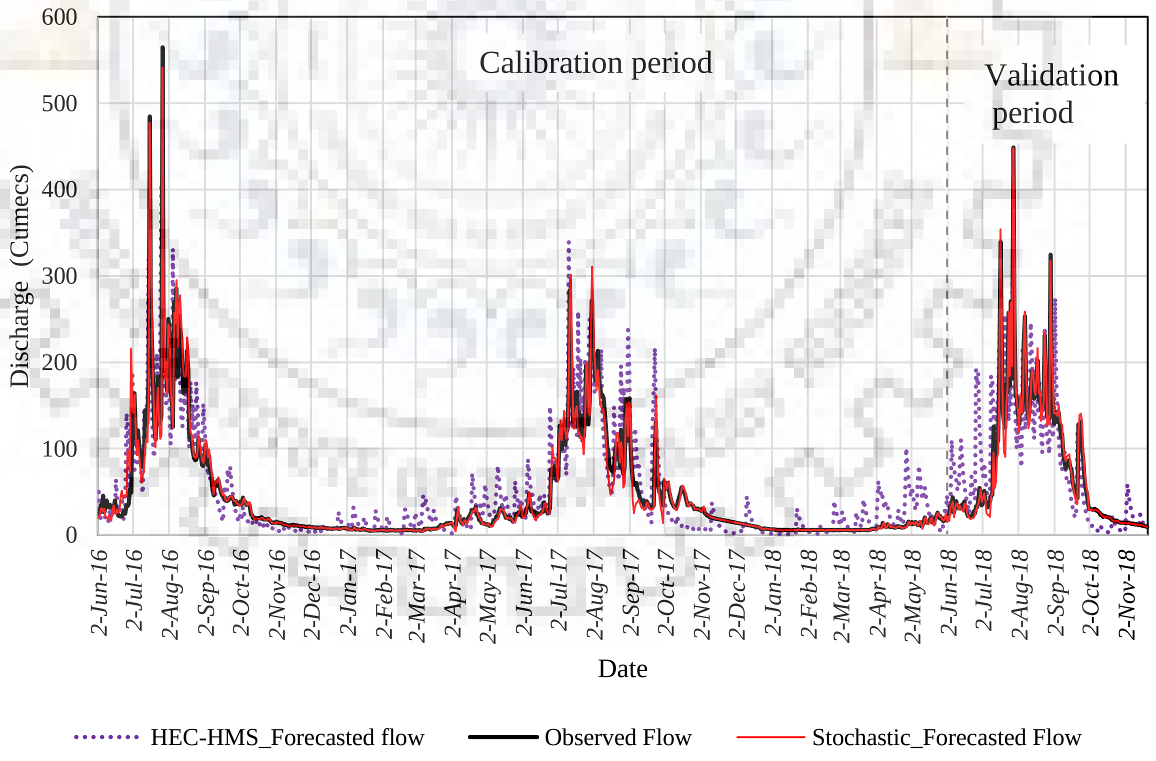


Figure 5.21. Observed and forecasted streamflow for Bhilangana sub-basins

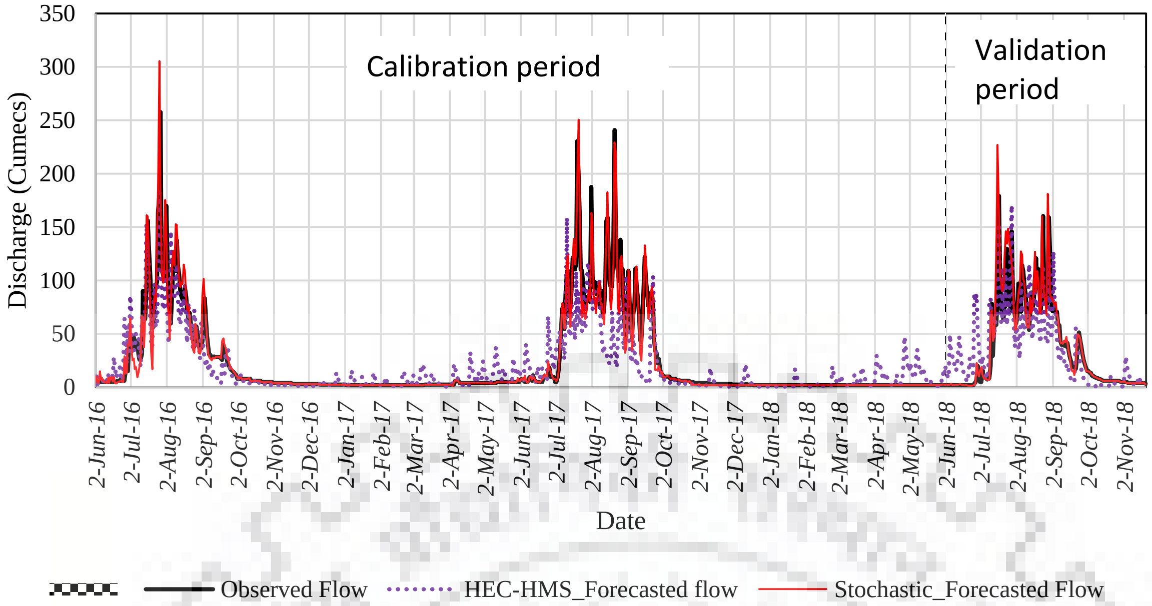


Figure 5.22. Observed and forecasted streamflow for Balganga sub-basins

5.6 RESULTS OF HEC-HMS MODEL OBTAINED IN THE PRESENT STUDY

5.6.1 Calibration of the Model

For the present study, the data of *1st June 2016 to 31st December 2017* were used for calibration of the model. The range of different parameter values used for the calibration purpose is given in Table 5.14. The calibrated parameter values for all four sub-catchments are given in Table 5.15-5.16.

Table 5.14.Maximum and minimum parameter values

Model	Parameter	Minimum Value	Maximum Value
SCS Loss	Initial Abstraction, Ia	0 mm	500 mm
	Curve Number, CN	1	100
Clark's UH	Storage Coefficient (R)	0 hr	150 hr
	Time of Concentration (Tc)	0.1 hr	500 hr
Base Flow	Initial Base Flow, Q_o	0 m^3/s	100000 m^3/s
	Recession Factor, R_c	0.000011	-
Muskingum Routing	K	0.1 hr	150 hr
	X	0	0.5
	Number of Steps	1	100

Table 5.15. Parameters values for all the sub-basins

Model	Parameter	Sub-basin	Value
SCS loss	Initial Abstraction	Bhilangana	3
		Balganga	5
		MB 2	5
		Tehri Dam	5
	Curve Number	Bhilangana	61
		Balganga	68
		MB 2	70
		Tehri Dam	60
Base flow	Initial Baseflow	Bhilangana	30
		Balganga	2
		MB 2	90
		Tehri Dam	60
	Recession Factor	Bhilangana	0.80
		Balganga	0.70
		MB 2	0.70
		Tehri Dam	0.75

Table 5.16. Calibrated parameters (except model component) for all the sub-basins.

Parameter	Sub-basin	Gage	Value
Gage Weights	Bhilangana	Bishan	0.35
		Dhoardhar	0.2
		Ghansali	0.45
	Balganga	Bishan	0.6
		Dhopardhar	0.1
		Ghansali	0.3
	MB 2	Bhatwari	0.1
		Dharasu	0.1
		Harshil	0.25
		Sukkhi	0.45
		Uttarkashi	0.1
	Tehri Dam	Dharasu	0.2
		Ghansali	0.4
		Lambgoan	0.2
		Tehri	0.2
Temperature Index	Bhilangana		20
	Balganga		22
	MB 2		20
	Tehri Dam		18

The calibration results for all the sub-catchments are given in Table 5.17. The NSE value is more than 76% for all the sub-basins, while the RMSE (Root Mean Square Error) values are also not very high. The lowest NSE value of 0.764 was obtained for Balganga sub-basin at the Sarasgaon gauging site, while the highest value of 0.798 was obtained for Bhagirathi basin at MB II. The highest RMSE value of 141.10 was found for the Bhagirathi basin at the Tehri dam.

Table 5.17. Observed and simulated results for calibrated daily runoff in all sub-basin.

Bhilangana Sub-basin at Ghansali		
Statistical Parameters	Observed	Simulated
Mean (m^3/sec.)	49.69	50.23
Standard deviation (m^3/sec.)	62.06	62.20
Maximum (m^3/sec.)	359.20	384.10
Minimum (m^3/sec.)	4.60	1.80
Nash-Sutcliffe Coefficient (E)	0.782	
Coefficient of Determination (r^2)	0.804	
Root Mean Square Error (m^3/sec.)	25.20	
Balganga Sub-basin at Sarasgaon		
Statistical Parameters	Observed	Simulated
Mean (m^3/sec.)	21.16	21.05
Standard deviation (m^3/sec.)	32.05	29.76
Maximum (m^3/sec.)	191.90	185.70
Minimum (m^3/sec.)	2.00	0.60
Nash-Sutcliffe Coefficient (E)	0.764	
Coefficient of Determination (r^2)	0.757	
Root Mean Square Error (m^3/sec.)	15.90	
Bhagirathi Sub-basin at MB II		
Statistical Parameters	Observed	Simulated
Mean (m^3/sec.)	158.39	159.03
Standard deviation (m^3/sec.)	177.18	165.82
Maximum (m^3/sec.)	1100.30	791.30
Minimum (m^3/sec.)	14.30	13.70
Nash-Sutcliffe Coefficient (E)	0.798	
Coefficient of Determination (r^2)	0.795	
Root Mean Square Error (m^3/sec.)	80.40	
Bhagirathi Sub-basin at Tehri Dam		
Statistical Parameters	Observed	Simulated
Mean (m^3/sec.)	271.29	218.12
Standard deviation (m^3/sec.)	304.63	278.32
Maximum (m^3/sec.)	1525.60	1761.60
Minimum (m^3/sec.)	12.70	16.90
Nash-Sutcliffe Coefficient (E)	0.789	
Coefficient of Determination (r^2)	0.823	
Root Mean Square Error (m^3/sec.)	140.08	

The comparison of observed and simulated daily runoff at all the gauging sites during the calibration periods are shown in Figure 5.23 to 5.26.

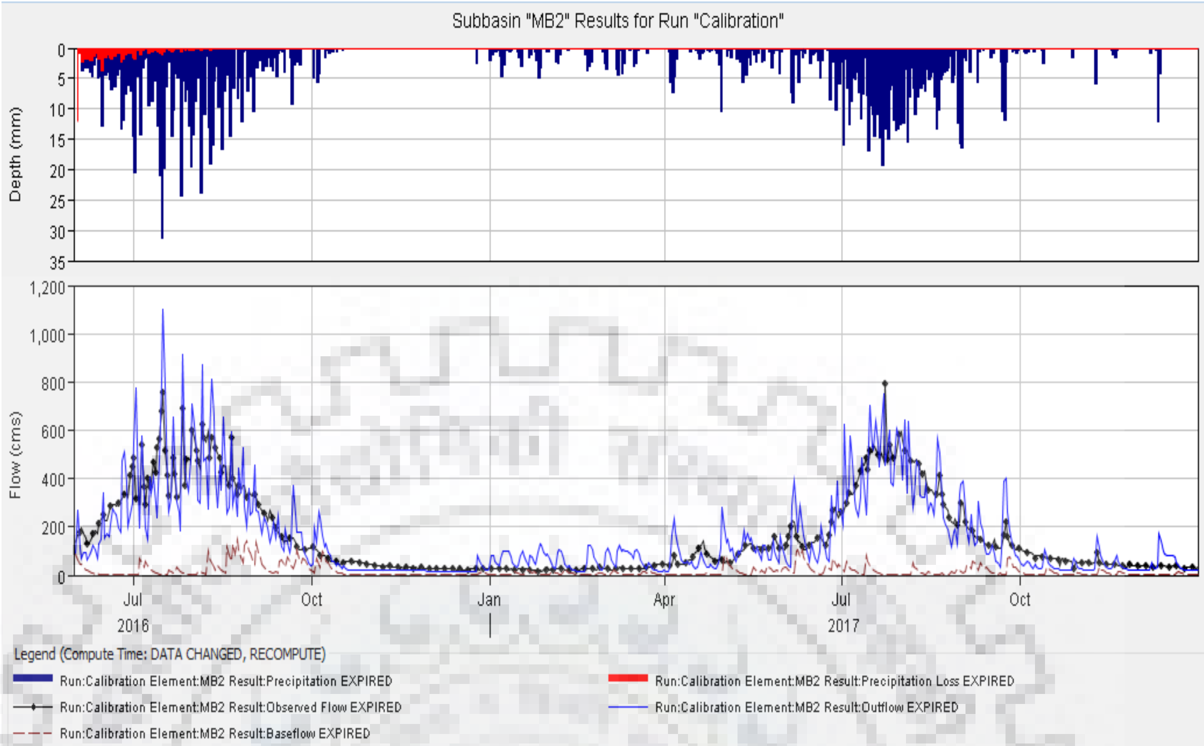


Figure 5.23.The simulated and observed runoff at Bhagirathi in MB II (calibration period).

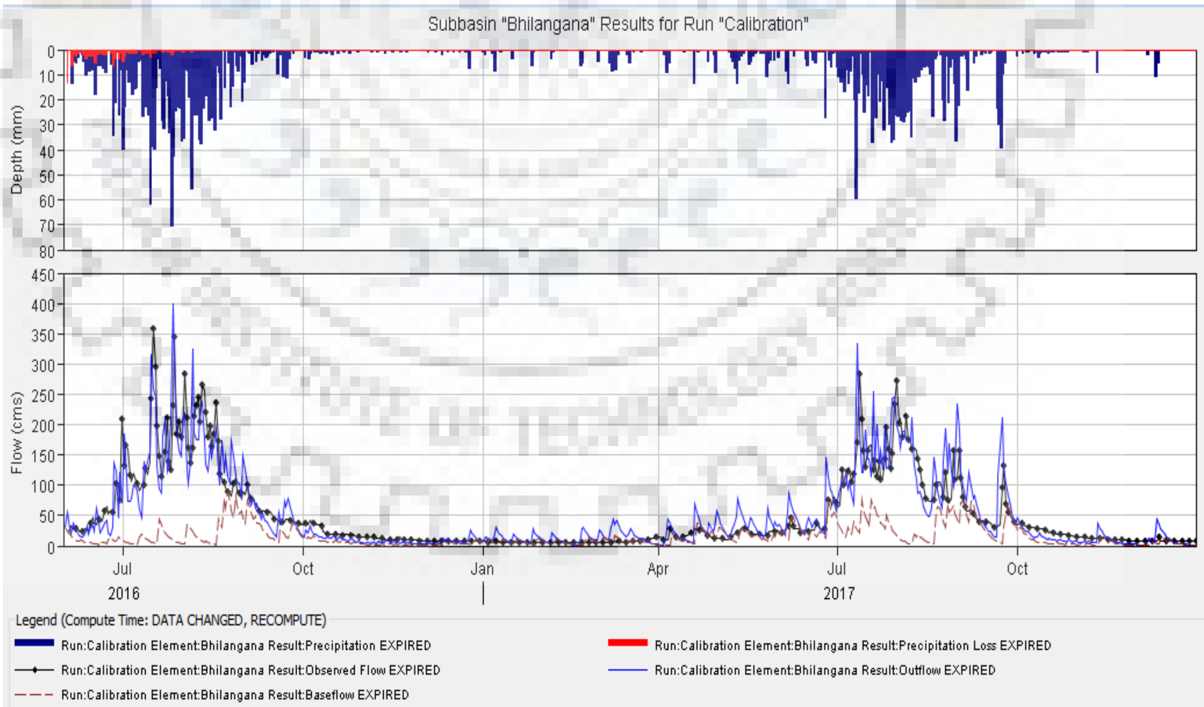


Figure 5.24. The simulated and observed runoff at Ghansali gauging site (calibration period).

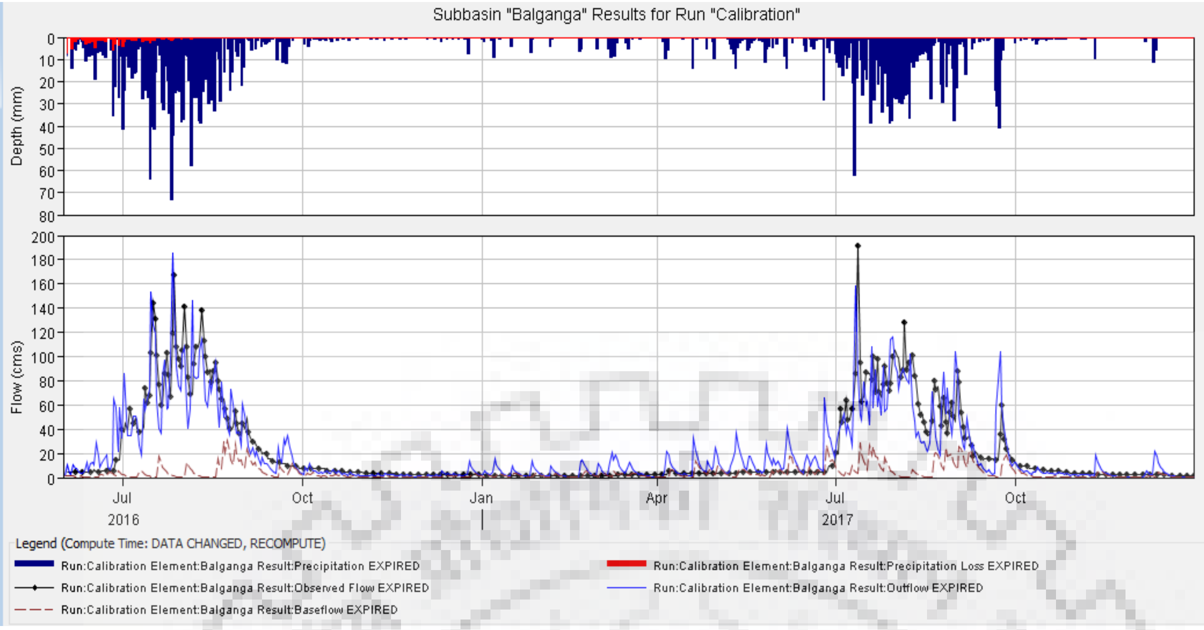


Figure 5.25. The simulated and observed runoff at Sarasgaon gauging site (calibration period).

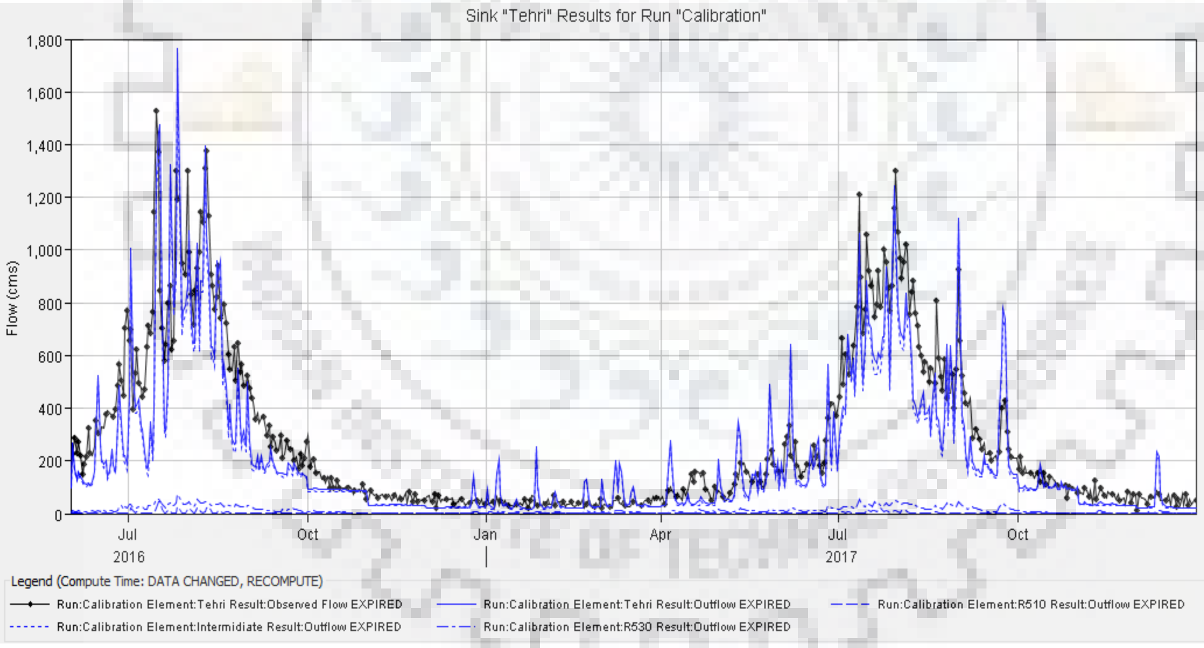


Figure 5.26. The simulated and observed runoff at Tehri Dam, in the calibration period.

5.6.2 Validation of the Model

In validation, the same calibrated parameters are used to check the model capability for simulating runoff. In the present study, the data from *1st January 2018 to 28th November 2018* were used for validation purpose. The results are given in Table 5.18. It can be observed that the NSE values for all the sub-basins are more than 70%. The comparison of observed and simulated runoff for all the sub-basins during validation are shown in Figure 5.27 to 5.30.

Table 5.18. Observed and simulated validation daily runoff for Bhagirathi river basin.

Bhilangana Sub-basin at Ghansali		
Statistical Parameters	Observed	Simulated
Mean (m^3/sec.)	43.51	43.55
Standard deviation (m^3/sec.)	64.52	62.22
Maximum (m^3/sec.)	448.60	351.50
Minimum (m^3/sec.)	5.31	0.20
Nash-Sutcliffe Coefficient (E)	0.768	
Coefficient of Determination (r^2)	0.790	
Root Mean Square Error (m^3/sec.)	31.40	
Balganga Sub-basin at Sarasgaon		
Statistical Parameters	Observed	Simulated
Mean (m^3/sec.)	18.65	20.23
Standard deviation (m^3/sec.)	33.02	34.17
Maximum (m^3/sec.)	179.10	179.10
Minimum (m^3/sec.)	2	2.00
Nash-Sutcliffe Coefficient (E)	0.708	
Coefficient of Determination (r^2)	0.705	
Root Mean Square Error (m^3/sec.)	18.50	
Bhagirathi Sub-basin at MB II		
Statistical Parameters	Observed	Simulated
Mean (m^3/sec.)	128.47	137.84
Standard deviation (m^3/sec.)	133.415	136.44
Maximum (m^3/sec.)	687.20	934.22
Minimum (m^3/sec.)	22.30	10.34
Nash-Sutcliffe Coefficient (E)	0.733	
Coefficient of Determination (r^2)	0.752	
Root Mean Square Error (m^3/sec.)	74.40	
Bhagirathi Sub-basin at Tehri Dam		
Statistical Parameters	Observed	Simulated
Mean (m^3/sec.)	221.937	218.80
Standard deviation (m^3/sec.)	279.508	287.74
Maximum (m^3/sec.)	1484.2	1647.13
Minimum (m^3/sec.)	15.20	12.70
Nash-Sutcliffe Coefficient (E)	0.723	
Coefficient of Determination (r^2)	0.756	
Root Mean Square Error (m^3/sec.)	132.20	

The results of the HEC-HMS application by Agrawal (2018) and the results obtained in this study clearly indicate that the performance of stochastic models is better than the HEC-HMS. Therefore, the stochastic models are chosen for daily inflow forecasting of Tehri catchment and the details are presented in next section.

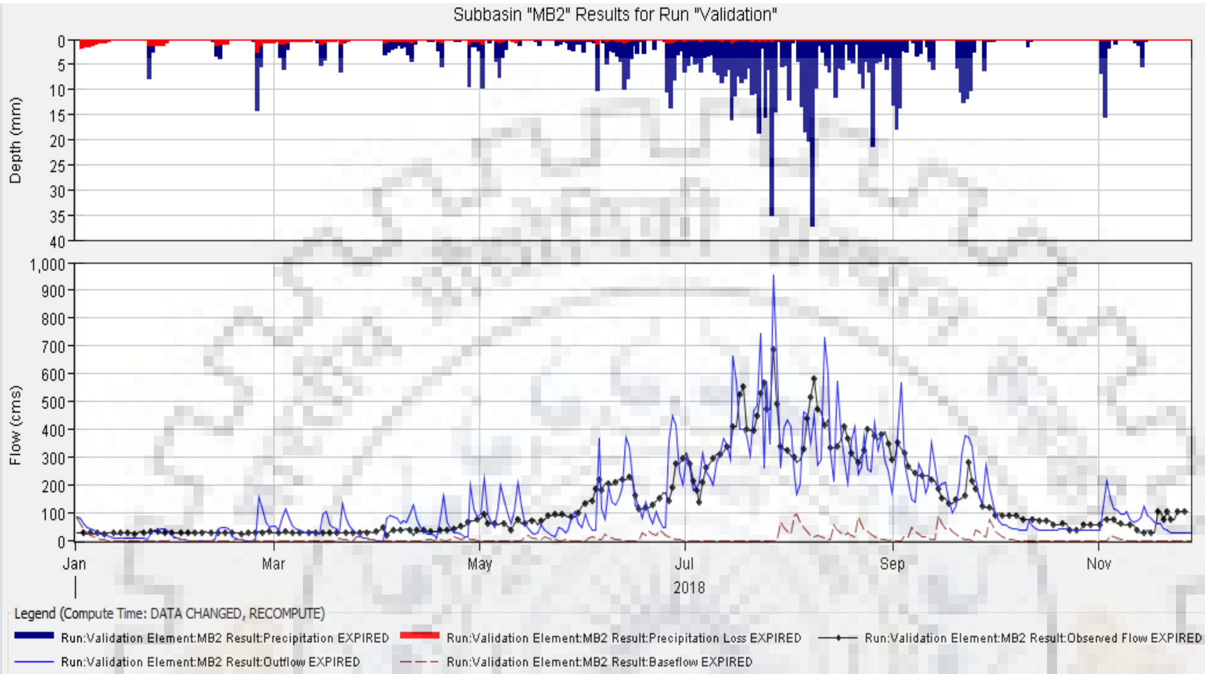


Figure 5.27. Plotted observed and simulated runoff at Bhagirathi in MB II validation period.

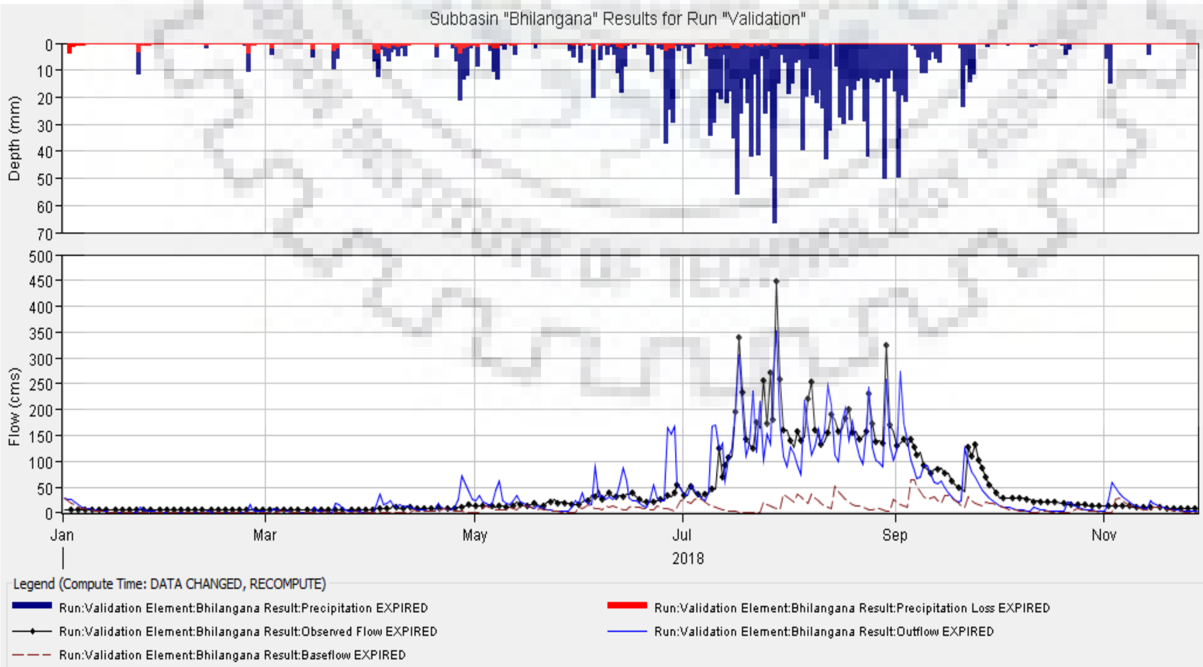


Figure 5.28. The simulated and observed runoff at Ghansali gauging site validation period.

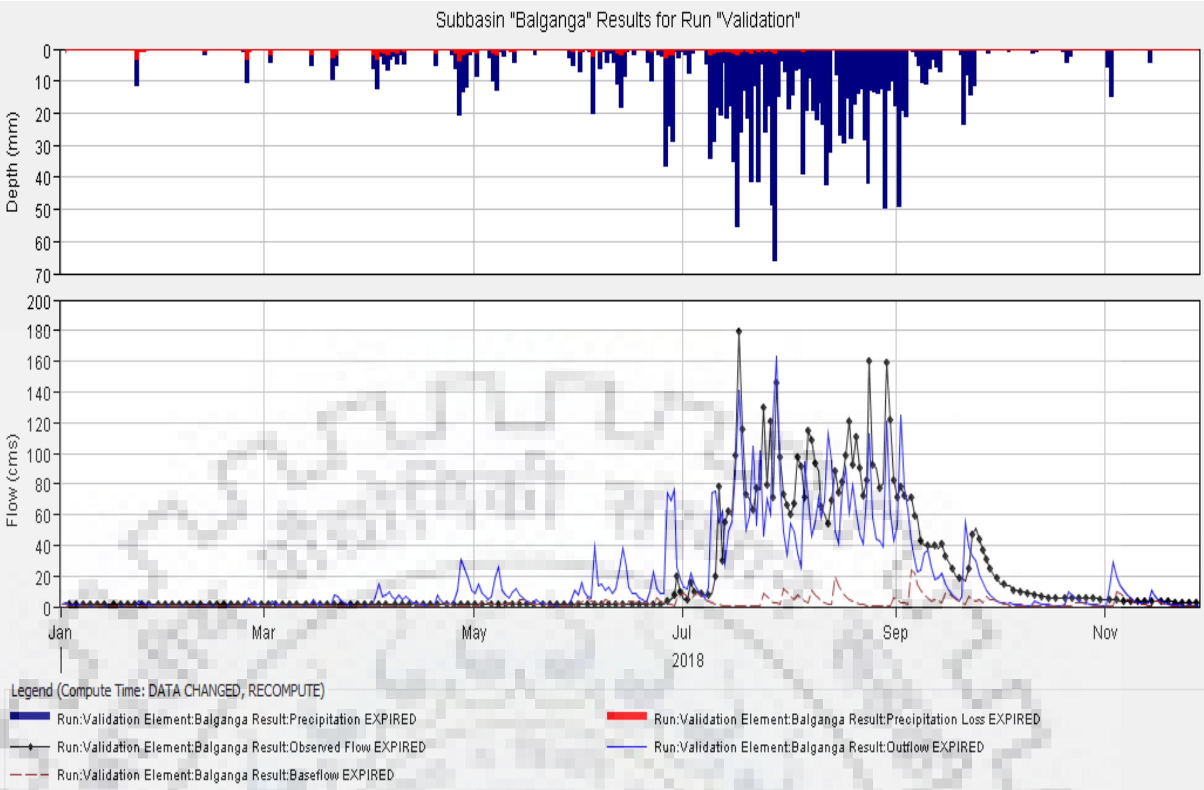


Figure 5.29. The simulated and observed runoff at Sarasgaon gauging site validation period.

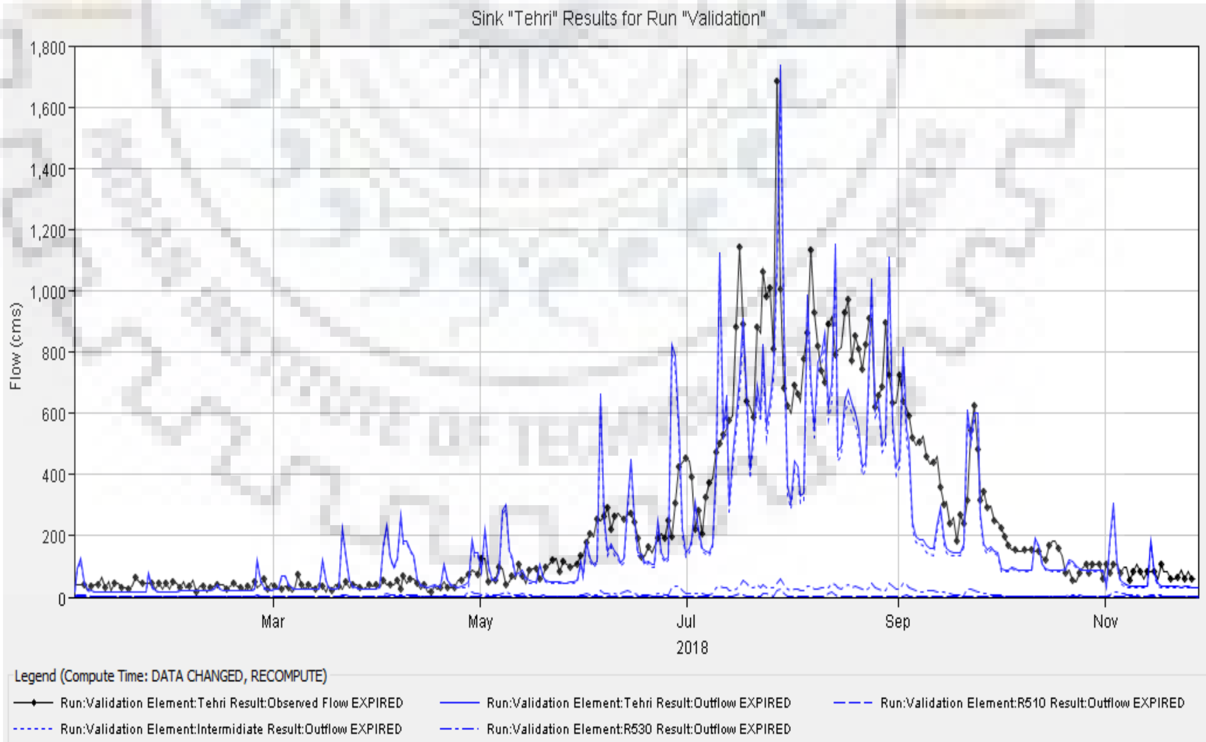


Figure 5.30. The simulated and observed runoff at Tehri Dam validation period.

5.7 FORECASTING OF DAILY RUNOFF USING STOCHASTIC MODELS

The best stochastic model has been used as a forecasting model to see the forecasting ability of the chosen stochastic model. During the forecasting, the model is simulated using the same model structure. However, for monsoon, the simulation period is taken from 1st June 2018 to 14th August 2018. The forecasting is done for the period of 15th August to 30th September 2018. For non-monsoon, the simulation period is taken from 1st October 2018 to 31st January 2019. The forecasting is done for the period of 1st February to 15th May 2019. The forecasting results for all the sub-basins are shown in Figure 5.31 to 5.36. The results clearly indicate the suitability of stochastic models for use in forecasting.

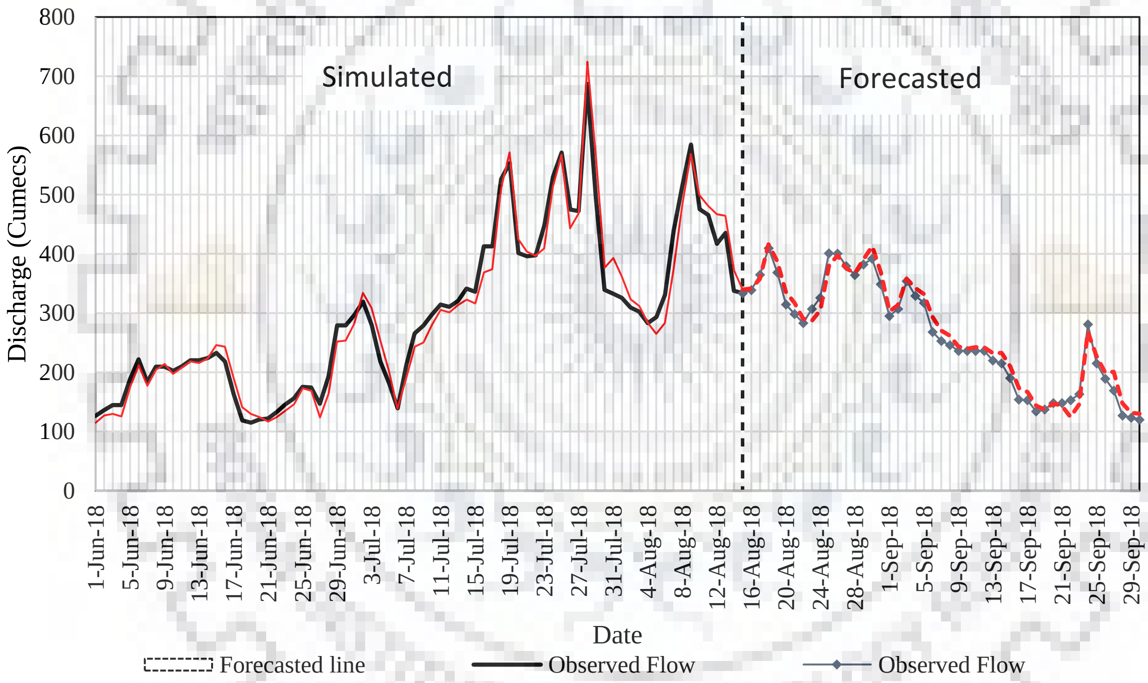


Figure 5.31. Forecasting streamflow for monsoon season at MB-II (Bhagirathi River Basin) by using Stochastic models (ARX(1,0,1))

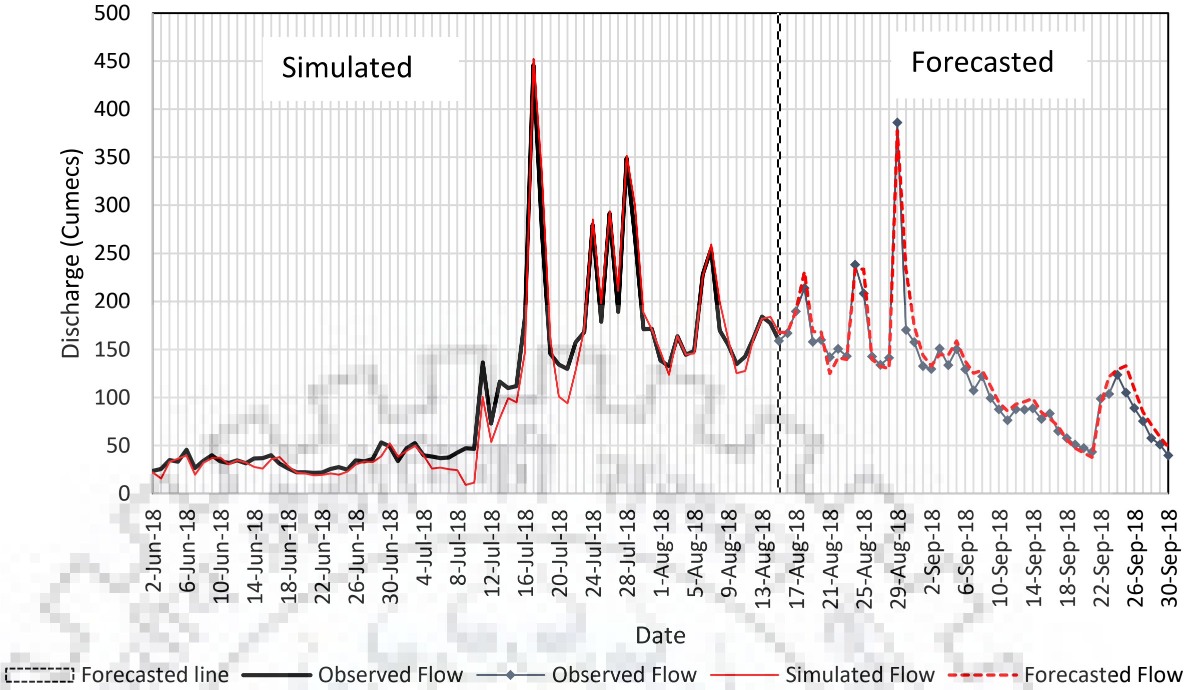


Figure 5.32. Forecasting streamflow for monsoon season at Ghansali gauging site by using stochastic models (ARMAX (1,1,1)).

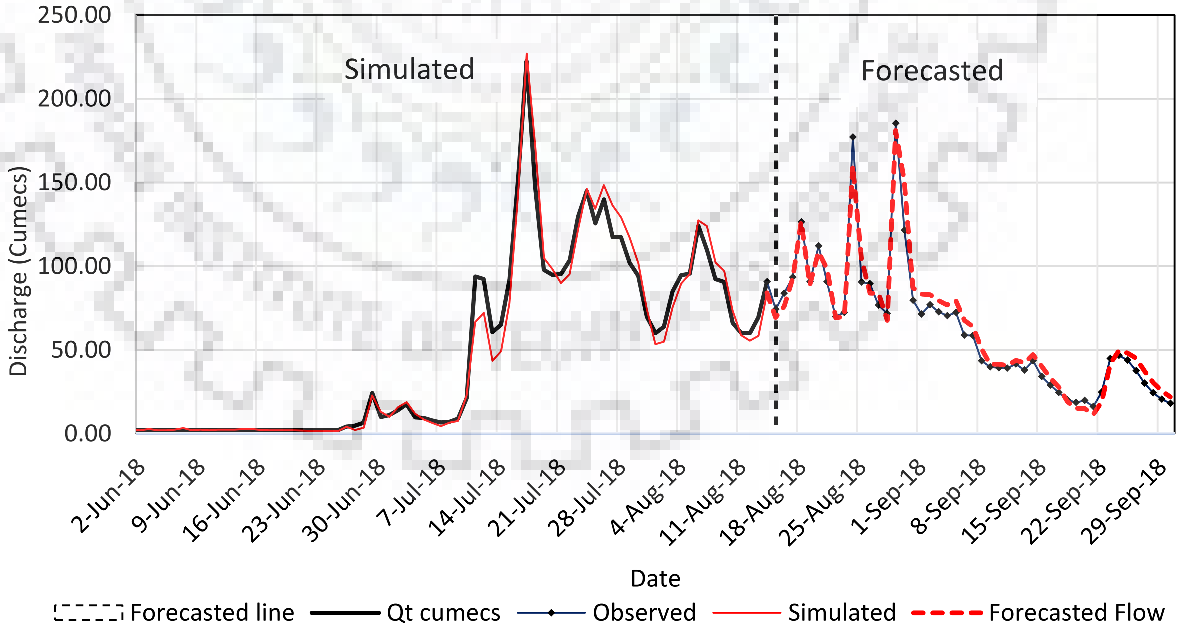


Figure 5.33. Forecasting streamflow for monsoon season at Sarasgaon gauging site by using Stochastic models (ARMAX (1,1,1)).

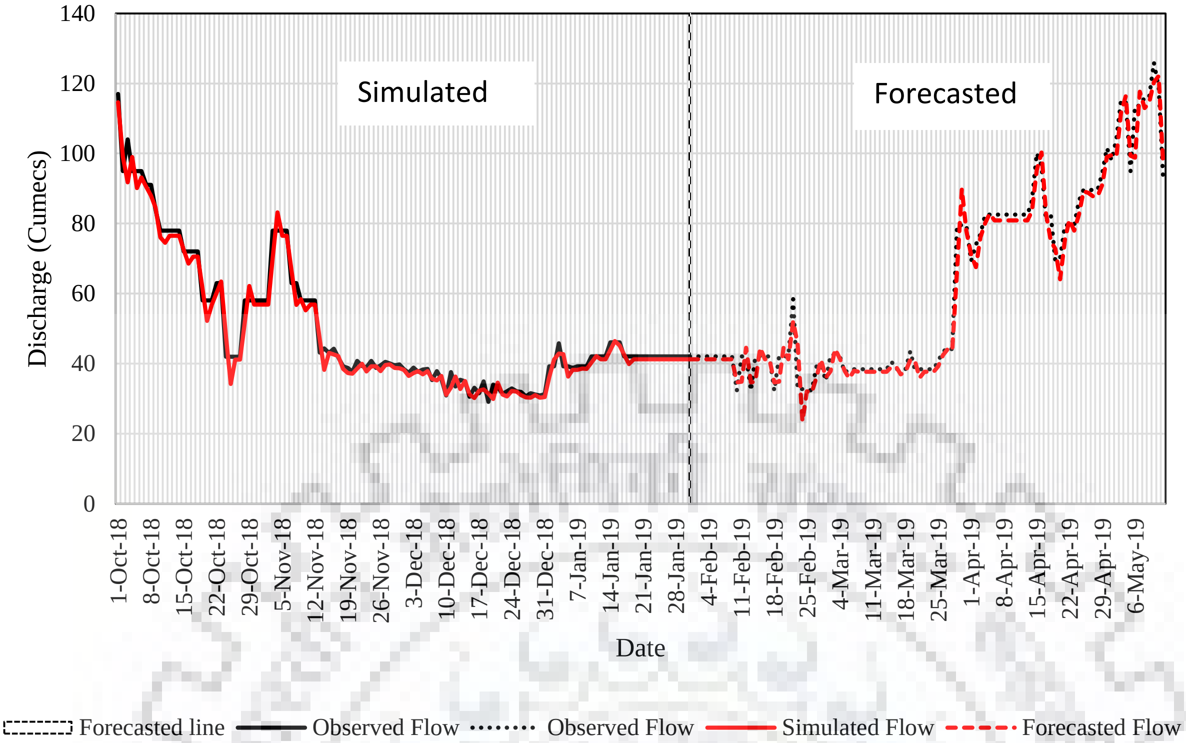


Figure 5.34. Forecasting streamflow for the non-monsoon season at MB-II (Bhagirathi River Basin) by using Stochastic models (AR(1))

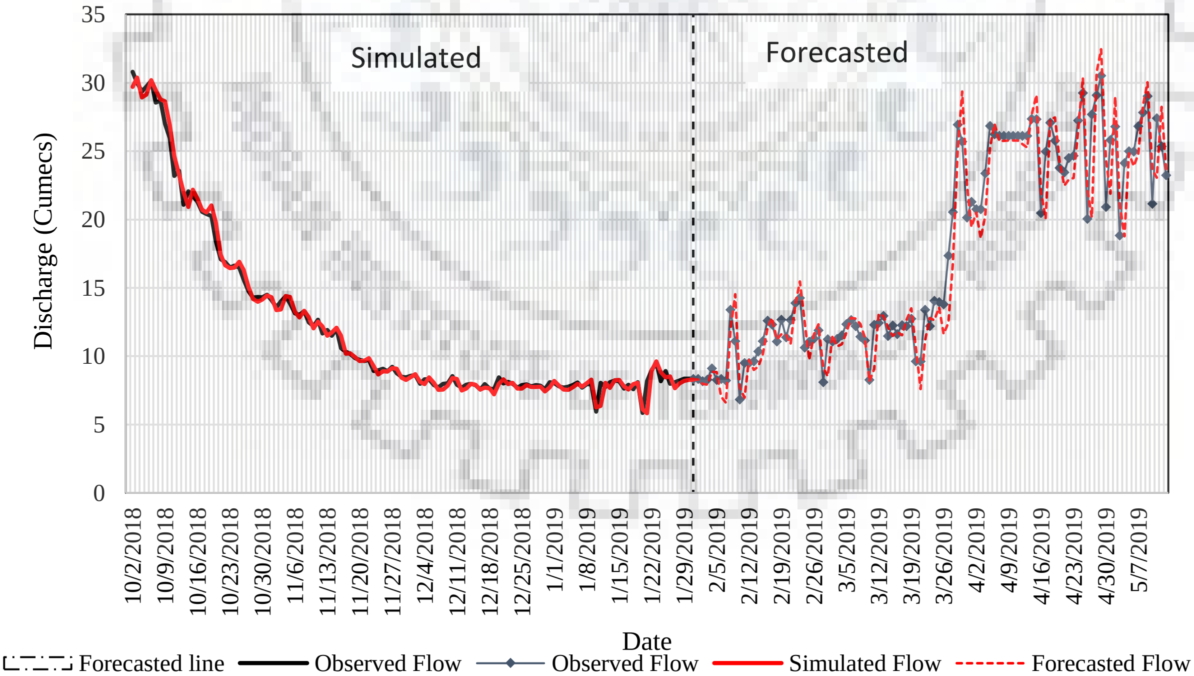


Figure 5.35. Forecasting streamflow for the non-monsoon season at Ghansali gauging site by using stochastic models (AR(1)).

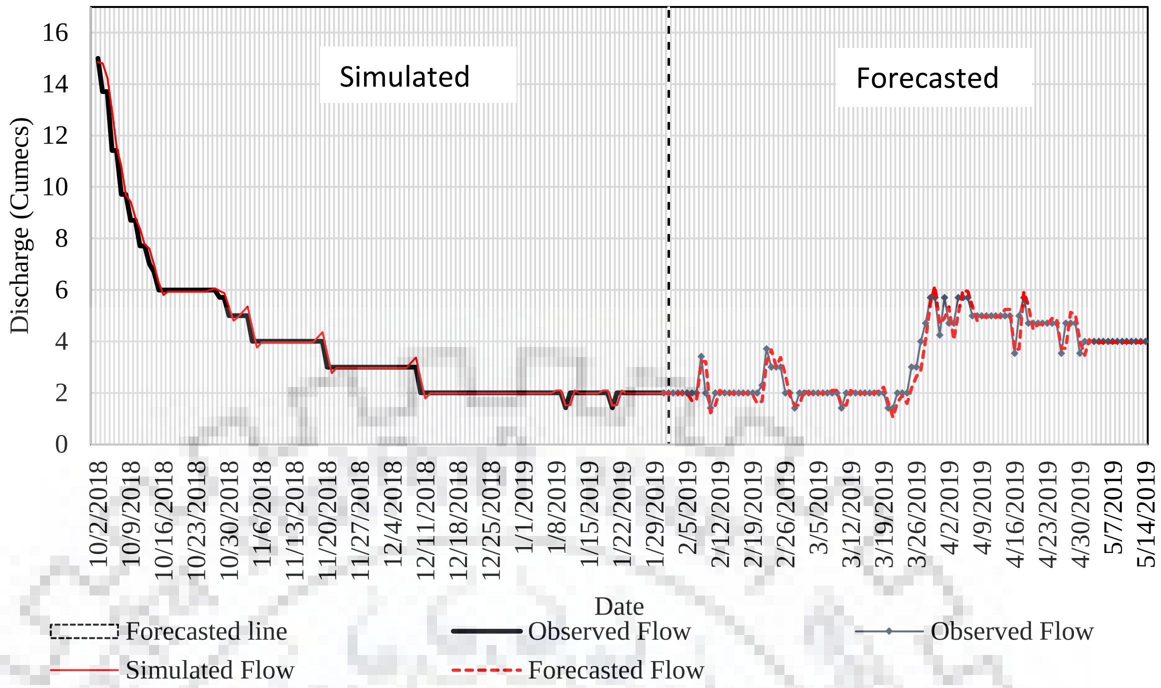


Figure 5.36. Forecasting streamflow for non-monsoon season at Sarasgaon gauging site by using Stochastic models (AR(1)).

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 GENERAL

Inflow forecasting of a storage dam is an important aspect due to its wider implication over society. Therefore, in the present study, an inflow forecasting system has been developed for Tehri reservoir. The aim is to improve the available information about the inflow to the Tehri reservoir which will give advance information and result in the improvement of the regulation of the spillway gates and the optimum generation of electricity for the full seasons (Monsoon and Non-monsoon). The system will play an essential role in disaster management of downstream of the reservoir. To fulfil the objective, at first, the rating curves have been developed for two sub-basins, namely Bhilangana and Balganga of Tehri catchment using method of least squares and ANN technique. Following this, the stochastic models have been developed for three main sub-catchments of Tehri dam. The results of the stochastic models have been compared with the results of HEC-HMS.

6.1.1 Development of the stage-discharge relationship

For developing the stage-discharge relationships, the data set of 1st June 2016 to 30th November 2018 from two gauging stations, namely Ghansali in Bhilangana river and Sarasgaon in Balganga river have been used. The performance of both the methods have been evaluated using Nash Sutcliffe Efficiency (NSE) and coefficient of determination (r^2). The following conclusions are drawn from the analysis of the data:

i. The results of the analysis show good performance by both the methods.
ii. For the method of least squares, the NSE was more than 95% and the coefficient of determination was more than 0.9. However, the efficiency of the ANN method was slightly better than the method of least squares. The RMSE was far less in the case of ANN.
iii. The equations developed using the method of least squares for the two sites are recommended to be used for the field application. The coefficients of these equations are in agreement with the physical analysis of cross sections of the two sites.

Bhilangna at Ghansali

$Q = 31.783*(H\text{-}Ho)^{1.900}$; $\mathbf{R^2 = 0.994}$

Range of applicability $849.40 \geq H \leq 855.00$

Balganga at Sarasgaon

$Q = 22.7743*(H\text{-}Ho)^{1.742}$; $\mathbf{R^2 = 0.997}$

Range of applicability $856.0 \geq H \leq 860$

6.1.2 Development of Stochastic models

Four stochastic models namely AR, ARX, ARMA and ARMAX have been developed for the three sites of the Tehri catchment. The rainfall and discharge data from June 2016 to May 15, 2019, for the three sub-basins, namely Bhagirathi at MB II, Bhilangana at Ghansali and Balganga at Sarasgaon were collected from Real-time inflow forecasting system website of Tehri dam. All the developed models were calibrated and validated by dividing the data into two parts. The performance of all the developed stochastic models has been checked using 6 indices namely NSE, RMSE, PBIAS%, R^2, MAE and AIC. The results of these models were compared with the results of HEC-HMS model. For the three sites, the models which performed the best on monsoon and non-monsoon basis during the calibration period are listed below:

(i) ARX (1,0,1) model gave an NSE of 0.986 and MAE of 15.9 Cumecs for Bhagirathi at MB II. This model is recommended for use in the monsoon period.

(ii) ARMAX (1,1,1) model gave an NSE of 0.953 and MAE of 15.127 Cumecs for Bhilangana at Ghansali. This model is recommended for use in the monsoon period.

(iii) ARMAX (1,1,1) model gave an NSE of 0.971 and MAE of 6.385 Cumecs for Balganga at Sarasgaon. This model is recommended for use in the monsoon period.

(iv) AR (1) model gave an NSE of 0.988 and MAE of 2.089 Cumecs for Bhagirathi at MB II. This model performed better than other models in terms of the six performance indicators used in the study. This model is recommended for use in the non-monsoon period.

(v) AR (1) model gave an NSE of 0.980 and MAE of 0.846 Cumecs for Bhilangana at Ghansali. This model is recommended for use in the non-monsoon period.

(vi) AR (1) model gave an NSE of 0.962 and MAE of 0.090 Cumecs for Balganga at Sarasgaon. This model performed better than other models in terms of the six performance indicators used in the study. This model is recommended for use in the non-monsoon period.

(vii) The comparison of Stochastic and HEC-HMS model shows that the performance of selected stochastic models is far better than the HEC-HMS model for the three sites of the Tehri catchment during calibration and validation both.

(viii) The forecasting ability of the stochastic model was also checked. The results confirm that the stochastic models can be used for the forecasting of daily streamflow of the three sites of the catchment.

(ix) The programs have also been prepared in R-studio version 3.4.3 software for the simulation of daily streamflow using stochastic models for all three sub-basins of the catchment.

6.2 RECOMMENDATIONS AND SCOPE FOR FURTHER WORK

The present study is the first step to develop an inflow forecasting system for Tehri dam using the data up to May 2019. Therefore, the present study could not be completed without limitations. The recommendations made on the basis of the study and scope for future work are given below:

- The stage-discharge relationship was drawn only using the data from 2016 to 2018, which may not cover the higher flood records and therefore, during the floods, the developed relationship may give lesser value than actual. For this, the relationship could be redrawn in future by using more dataset and a new relationship can be drawn only for flood situation i.e. for higher values of the flood stages.
- In case of the stochastic model, only AR model was developed for non-monsoon season. In future, development of other stochastic models considering the rainfall and temperature are expected to give better results.
- More efforts are required to be put in for increasing the efficiency of the HEC-HMS model with extended data bases. With extended data base, the efficiency of HEC-HMS is expected to improve further.
- The updating of parameters of stochastic models on a daily basis is recommended in future work.

REFERENCES

Adamowski, J., & Adamowski, K. (2008). Development of a Real-time River Flood Forecastinig Transfer Function-Noise Model with a Kalman Filter for Snowmelt Driven Floods. *J. Environmental Hydrology, 16*(August), 1-11.

Adnan, R., Yuan, X., Kisi, O., & Yuan, Y. (2017). Streamflow Forecasting Using Artificial Neural Network and Support Vector Machine Model. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS), 29*(1), 286-294.

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(5), 717-727.

Agrawal, N. K. (2018). Development of an Operational Inflow forecasting system for Tehri dam. *PhD thesis.* Department of Hydrology, Indian Institute of Technology Roorkee: Unpublished Work, 1-133.

Aichouri, I., Hani, A., Bougherira, N., Djabri, L., Chaffai, H., & Lallahem, S. (2015). River Flow Model Using Artificial Neural Networks. *Energy Procedia, 74*, 1007-1014.

Akouemo, H., & Povinelli, R. (2017). Data Improving in Time Series Using ARX and ANN Models. *IEEE Transactions on Power Systems, 32*(5), 3352-3359.

Akshaykumar, N., & Subbulekshmi, D. (2016). Process identification with autoregressive linear regression method using experimental data: Review. *Indian Journal of Science and Technology, 9*(39), 1-7.

Alam, M. (2016). Codes in MATLAB for Training Artificial Neural Network using Particle Swarm Optimization. (August), 1-16.

Alfa, M., Adie, D., Ajibike, M., & Mudiare, O. (2018). Development of rating curve for Ofu River at Oforachi Hydrometric Station. *Nigerian Journal of Technological Development*, *15*(1), 1-14.

Asadi, A. (2013). Application of HEC-HMS for Flood Forecasting in Kabkian Basin and Delibajak Subbasin in Iran. *IOSR Journal of Engineering, 03*(09), 10-16.

ASCE Committee. (2000). Task Committee on Application of Artificial Neural Networks in Hydrology, Artificial Neural Networks in Hydrology. II:Hydrologic Application. *Journal of Hydrologic Engineering, 5*(2), 124-136.

Awchi, T. (2014). River discharges forecasting in northern Iraq using different ANN techniques. *Water Resources Management, 28*(3), 801-814.

Bauer-Gottwein, P., Jensen, I., Guzinski, R., Bredtoft, G., Hansen, S., & Michailovsky, C. (2015). Operational river discharge forecasting in poorly gauged basins: The Kavango River basin case study. *Hydrology and Earth System Sciences, 19*(3), 1469-1485.

Chibanga, R., Berlamont, J., & Vandewalle, J. (2003). Modelling and forecasting of hydrological variables using artificial neural networks: The Kafue River sub-basin. *Hydrological Sciences Journal, 48*(3), 363-380.

Choudhari, K., Panigrahi, B., & Paul, J. (2014). Simulation of rainfall-runoff process using HEC-HMS model for Balijore Nala watershed, Odisha, India. *INTERNATIONAL Journal Of Geomatics And Geosciences, 5(2),* 253-265.

Chow, V. (2010). Bibliography: 1) Handbook of Applied Hydrology. *International Association of Scientific Hydrology. Bulletin, 10*(1), 82-83.

Cigizoglu, H. (2005). Application of Generalized Regression Neural Networks to Intermittent Flow Forecasting and Estimation. *Journal of Hydrologic Engineering, 10*(4), 336-341.

Collischonn, W., Haas, R., Andreolli, I., & Tucci, C. (2005). Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. *Journal of Hydrology, 305*(1-4), 87-98.

Costa, J., Spicer, K., Cheng, R., Haeni, F., Melcher, N., Thurman, E., & Keller, W. (2000). Measuring stream discharge by non-contact methods: A proof-of-concept experiment. *Geophysical Research Letters, 27*(4), 553-556.

Dashora, I., Singal, S. K., & Srivastav, D. K. (2015). Software Application for Data Driven Prediction Models for Intermittent Streamflow for Narmada River Basin. *International Journal of Computer Applications, 113*(10), 9-17.

Dolling, O., & Varas, E. (2002). Utilisation des réseaux des neurones artificielles pour la prédiction des écoulements. *Journal of Hydraulic Research, 40*(5), 547-554.

Durdu, Ö. (2010). Stochastic approaches for time series forecasting of boron: A case study of Western Turkey. *Environmental Monitoring and Assessment, 169*(1-4), 687-701.

Dutta, D., Wilson, K., Welsh, W., Nicholls, D., Kim, S., & Cetin, L. (2013). A new river system modelling tool for sustainable operational management of water resources. *Journal of Environmental Management, 12*(1), 13-28.

Elganiny, M., & Eldwer, A. (2016). Comparison of Stochastic Models in Forecasting Monthly Streamflow in Rivers: A Case Study of River Nile and Its Tributaries. *Journal of Water Resource and Protection, 08*(02), 143-153.

Elganainy, M., & Eldwer, A. (2018). Stochastic Forecasting Models of the Monthly Streamflow for the Blue Nile at Eldiem Station. *Water Resources, 45*(3), 326-337.

El-Shafie, A., Abdin, A., Noureldin, A., & Taha, M. (2009). Enhancing inflow forecasting model at aswan high dam utilizing radial basis neural network and upstream monitoring stations measurements. *Water Resources Management, 23*(11), 2289-2315.

Guven, A., & Aytek, A. (2009). New Approach for Stage–Discharge Relationship: Gene-Expression Programming. *Journal of Hydrologic Engineering, 14*(8), 812-820.

Haan, C., Barfield, B., & Hayes, J. (2012). Hydrologic Modeling. *Design Hydrology and Sedimentology for Small Catchments, 72*(2), 455-474.

Haltiner, J., & Salas, J. (1988). SHORT-TERM FORECASTING OF SNOWMELT RUNOFF USING ARMAX MODELS. *JAWRA Journal of the American Water Resources Association, 24*(5), 1083-1089.

Hasanah, Y., Herlina, M., & Zaikarina, H. (2013). Flood Prediction using Transfer Function Model of Rainfall and Water Discharge Approach in Katulampa Dam. *Procedia Environmental Sciences, 17*(1), 317-326.

Hassan, S., & Khan, H. (2015). Stochastic River Flow Modelling and Forecasting of Upper Indus Basin. *Journal of Basic & Applied Sciences, 11*(12), 630-636.

Hsu, K., Moradkhani, H., & Sorooshian, S. (2009). A sequential Bayesian approach for hydrologic model selection and prediction. *Water Resources Research, 45*(1), 1-15.

Kuo-lin, H., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process that arise and based Background and Scope. *Water Resources, 31*(10), 2517-2530.

Joorabchi, A., Zhang, H., & Blumenstein, M. (2007). Application of artificial neural networks in flow discharge prediction for the Fitzroy River , Australia. *2007*(50), 287-291.

Karl, A., & Lohani, A. (2010). Development of Flood Forecasting System Using Statistical and ANN Techniques in the Downstream Catchment of Mahanadi Basin, India. *Journal of Water Resource and Protection*, *02*(10), 880-887.

Kim, M., & Gilley, J. (2008). Artificial Neural Network estimation of soil erosion and nutrient concentrations in runoff from land application areas. *Computers and Electronics in Agriculture*, *64*(2), 268-275.

Kişi, Ö. (2005). Daily river flow forecasting using artificial neural networks and auto-regressive models. *Turkish Journal of Engineering and Environmental Sciences*, *29*(1), 9-20.

Kişi, Ö. (2007). Streamflow Forecasting Using Different Artificial Neural Network Algorithms. *Journal of Hydrologic Engineering*, *12*(5), 532-539.

Kisi, O., Shiri, J., & Tombul, M. (2013). Modeling rainfall-runoff process using soft computing techniques. *Computers and Geosciences*, *51*, 108-117.

Kuo, J., Hsieh, M., Lung, W., & She, N. (2007). Using artificial neural network for reservoir eutrophication prediction. *Ecological Modelling*, *200*(1-2), 171-177.

Lohani, A., Goel, N., & Bhatia, K. (2006). Takagi-Sugeno fuzzy inference system for modeling stage-discharge relationship. *Journal of Hydrology*, *331*(1-2), 146-160.

Lohani, A., Kumar, R., & Singh, R. (2012). Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology*, *442-443*, 23-35.

Majidi, A., & Shahedi, K. (2012). Simulation of Rainfall-Runoff Process Using Green-Ampt Method and HEC-HMS Model (Case Study: Abnama Watershed, Iran). *International Journal of Hydraulic Engineering*, *1*(1), 5-9.

McCuen, R. (2003). Time series modelling of water resources and environmental systems. *Journal of Hydrology*, *167*(1-4), 399-400.

Mishra, A., & Desai, V. (2005). Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment*, *19*(5), 326-339.

Poonia, V. (2018). Hydrological Analysis by Artificial Neural Network : A Review. *International Journal of Advance Research, Ideas and Innovations in Technology*, *4*(3), 265-270.

Salas, J. D., Delleur, J. W., Yevjevich, V., & Lane, W. L. (1980). APPLIED MODELLING OF HYDROLOGIC TIME SERIES. Chelsea, Michigan, U.S.A: Water Resources Publications, 1-484.

Salas, J. D., & Wang, D. C. (1991). FORECASTING STREAMFLOW FOR COLORADO RIVER SYSTEMS. Colorado: Fort Collins, 1-245.

Sardoii, E., Rostami, N., Sigaroudi, S., & Taheri, S. (2012). Calibration of loss estimation methods in HEC-HMS for simulation of surface runoff (case study: Amirkabir dam watershed, Iran). *Advances in Environmental Biology*, *6*(1), 343-348.

Sarhadi, A., Kelly, R., & Modarres, R. (2014). Snow water equivalent time-series forecasting in Ontario, Canada, in link to large atmospheric circulations. *Hydrological Processes*, *28*(16), 4640-4653.

Scharffenberg, W. (2013). Hydrologic Modeling System HEC-HMS, User Manual. (December), 1-442.

Sivapragasam, C., Vanitha, S., Muttil, N., Suganya, K., Suji, S., Thamarai Selvi, M., & Jeya Sudha, S. (2014). Monthly flow forecast for Mississippi River basin using artificial neural networks. *Neural Computing and Applications*, *24*(7-8), 1785-1793.

Upadhyay, T., Sankhayan, P., & Solberg, B. (2005). A review of carbon sequestration dynamics in the Himalayan region as a function of land-use change and forest/soil degradation with special reference to Nepal. *Agriculture, Ecosystems and Environment*, *105*(3), 449-465.

Wu, C., Chau, K., & Li, Y. (2009). Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *45*(May), 1-23.

Yadav, R., Kalra, P., & John, J. (2007). Time series prediction with single multiplicative neuron model. *Applied Soft Computing Journal*, *7*(4), 1157-1163.

Zgü, O., & Kisi, K. (2004). River Flow Modeling Using Artificial Neural Networks. *9*(February), 60-63.

Zhang, Z., Zhang, Q., & Singh, V. (2018). Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. *Hydrological Sciences Journal, 63*(7), 1091-1111.

APPENDIX-I

R SCRIPT PROGRAM DEVELOPED AND USED IN THE STUDY

##

R studio programming language have used for calibration and validation in development of the AR, ARX and ARMA models to forecasting daily streamflow for Tehri sub-basin.

DESCRIPTION CALIBRATION FOR THE TIME SERIES ANALYSIS.

The datasets used in the program is Discharge and Rainfall. Date, discharge and rainfall are prepared in a spreadsheet and exported as CSV ("comma-separated value") file named 'SARA_ARMAX-2016.CSV'.

1. Setwd ("E:/WORK DIS/R/Sarasgaon")   #####load the directory file by using setwd

2. Data = read.csv ('SARA_ARMAX-2016.CSV', TRUE, ",")

```
attach(Data)
class(Data)
head(Data)
```

Selecting and Run

3. ###### load R packages in the library

```
library(ggplot2)       # Creat Elagant Data Visulisations Using Grammar of Graphics
library(MASS)          # Support Functions and Dataset for venables and Ripley's MASS
library(tseries)       # Time series Analysis and Computational Finance
library(forecast)      # Forecasting Functions for Time series and Linera Models
library(tidyverse)     # data manipulation and visualization
library(lubridate)     # easily work with dates and times
library(fpp2)          # Data for "Forecasting: Principles and Practice" (2nd Edition)
library(zoo)           # S3 Infrastructure for Regular and Irregular Time series
library(dplyr)         # A Grammar data Manipulation
library(scales)        # Scale Functions for Visualization
library(quantmod)      # Quantitative Financial Modelling Framework
library(readr)         # Read reactangular Text data
```

4. ###### Data preparation with lag

```
Q <- (Flow [4:122])
Q_1 <- (Flow [3:121])
Q_2 <- (Flow [2:120])
Q_3 <- (Flow [1:119])
R   <- (Rainfall [4:122])
R_1 <- (Rainfall [3:121])
R_2 <- (Rainfall [2:120])
R_3 <- (Rainfall [1:119])
D <- (Date [4:122])
```

Selecting and Run

5. ###### Plotting the dataset using ggplot2 or normal graph plot

```
newdate <- as. Date (Data$Date, "%m/%d/%Y")
ggplot (Data, aes (newdate, Flow)) + geom_line (colour = "Blue") + scale_x_date (labels = date_format ("%b-%Y"), limits = c (as. Date ("2016-06-01"), as. Date ("2016-10-2"))) + ylab ("Discharge in Cumecs") + xlab("Date")

plot (as. Date (newdate, "%d-%b-%y"), Data$Flow, xlab = "Dates", ylab = "Discharge in Cumecs", type = "l", col = "red", main = "Balganga River_Flow 2016"
```

Selecting and Run

6. ###### Computing error

```
MA <- rollmean (Flow, 5)     #Error
F_1 <- Flow [3:122]
E1   <- (F_1-MA)
E_1 <- (E1[1:119])
E_t <- (E1[2:120])
E   <-  E_t
```

7. ###### Computing parameters (1,0,0), (2,0,0), (3,0,0), (1,1,0), (2,1,0), (3,1,0), (1,1,1), (2,1,1), (3,1,1), (1,1,1) by using matrix form.

Data transpose and multiply data matrix (1,0,0)

```
dim(Q_1) = c(119,1)
Qd = Q_1
Qt = t(Q_1)
```

```
        M1 = Qt%*%Qd
###### Inverse data
        B1 = solve(M1)
###### inverse multiply Data transpose
        C1 = B1%*%Qt
###### Parameter of (1,0,0)
        dim(Q) = c(119,1)
        Qobs = Q
        P1= C1%*%Qobs
        P1

###### Data transpose and multiply data matrix (2,0,0)
        Q2 = cbind (Q_1, Q_2)
        dim(Q2) = c(119,2)
        Qd2 = Q2
        Qt2 = t(Q2)
        M2 = Qt2%*%Qd2
###### Inverse data
        B2 = solve(M2)
###### inverse multiply Data transpose
        C2 = B2%*%Qt2
###### Parameter of (2,0,0)
        dim(Q) = c(119,1)
        Qobs = Q
        P2= C2%*%Qobs
        P2
###### Data transpose and multiply data matrix (3,0,0)
        Q3 = cbind(Q_1, Q_2, Q_3)
        dim(Q3) = c(119,3)
        Qd3 = Q3
        Qt3 = t(Q3)
        M3 = Qt3%*%Qd3
###### Inverse data
        B3 = solve(M3)
```

```
###### inverse multiply Data transpose
        C3 = B3%*%Qt3
###### Parameter of (3,0,0)
        dim(Q) = c(119,1)
        Qobs = Q
        P3= C3%*%Qobs
        P3
###### Data transpose and multiply data matrix (1,1,0)
        QE2 = cbind (Q_1, E_1)
        dim(QE2) = c(119,2)
        QEd2 = QE2
        QEt2 = t(QE2)
        ME2 = QEt2%*%QEd2
###### Inverse data
        BE2 = solve(ME2)
###### inverse multiply Data transpose
        CE2 = BE2%*%QEt2
###### Parameter of (1,1,0)
        dim(Q) = c(119,1)
        Qobs = Q
        PE2= CE2%*%Qobs
        PE2

###### Data transpose and multiply data matrix (2,1,0)
        QE3 = cbind(Q_1, Q_2, E_1)
        dim(QE3) = c(119,3)
        QEd3 = QE3
        QEt3 = t(QE3)
        ME3 = QEt3%*%QEd3
###### Inverse data
        BE3 = solve(ME3)
###### inverse multiply Data transpose
        CE3 = BE3%*%QEt3
###### Parameters of (2,1,0)
```

```
        dim(Q) = c(119,1)
        Qobs = Q
        PE3= CE3%*%Qobs
        PE3

###### Data transpose and multiply data matrix (3,1,0)
        QE4 = cbind(Q_1, Q_2, Q_3, E_1)
        dim(QE4) = c(119,4)
        QEd4 = QE4
        QEt4 = t(QE4)
        ME4 = QEt4%*%QEd4
###### Inverse data
        BE4 = solve(ME4)
###### inverse multiply Data transpose
        CE4 = BE4%*%QEt4
###### Parameters of (3,1,0)
        dim(Q) = c(119,1)
        Qobs = Q
        PE4= CE4%*%Qobs
        PE4

###### Data transpose and multiply data matrix (1,1,1)
        QER3 = cbind(Q_1, E_1, R_1)
        dim(QER3) = c(119,3)
        QERd3 = QER3
        QERt3 = t(QER3)
        MER3 = QERt3%*%QERd3
###### Inverse data
        BER3 = solve(MER3)
###### inverse multiply Data transpose
        CER3 = BER3%*%QERt3
##### Parameters of (1,1,1)
        dim(Q) = c(119,1)
        Qobs = Q
```

```
    PER3= CER3%*%Qobs
    PER3

###### Data transpose and multiply data matrix (2,1,1)
    QER4 = cbind(Q_1, Q_2, E_1, R_1)
    dim(QER4) = c(119,4)
    QERd4 = QER4
    QERt4 = t(QER4)
    MER4 = QERt4%*%QERd4
###### Inverse data
    BER4 = solve(MER4)
###### inverse multiply Data transpose
    CER4 = BER4%*%QERt4
###### Parameters of (2,1,1)
    dim(Q) = c(119,1)
    Qobs = Q
    PER4= CER4%*%Qobs
    PER4

###### Data transpose and multiply data matrix (3,1,1)
    QER5 = cbind(Q_1, Q_2, Q_3, E_1, R_1)
    dim(QER5) = c(119,5)
    QERd5 = QER5
    QERt5 = t(QER5)
    MER5 = QERt5%*%QERd5
###### Inverse data
    BER5 = solve(MER5)
###### inverse multiply Data transpose
  CER5 = BER5%*%QERt5
###### Parameters of (3,1,1)
    dim(Q) = c(119,1)
    Qobs = Q
    PER5 = CER5%*%Qobs
    PER5
```

```
###### Data transpose and multiply data matrix (2,0,1)
        QR3 = cbind(Q_1, Q_2, R_1)
        dim(QR3) = c(119,3)
        QRd3 = QR3
        QRt3 = t(QR3)
        MR3 = QRt3%*%QRd3
###### Inverse data
        BR3 = solve(MR3)
###### inverse multiply Data transpose
        CR3 = BR3%*%QRt3
###### Parameters of (2,0,1)
        dim(Q) = c(119,1)
        Qobs = Q
        PR3 = CR3%*%Qobs
        PR3

###### Data transpose and multiply data matrix (1,1,2)
        QER_4 = cbind(Q_1, E_1, R_1, R_2)
        dim(QER_4) = c(119,4)
        QERd_4 = QER_4
        QERt_4 = t(QER_4)
        MER_4 = QERt_4%*%QERd_4
###### Inverse data
        BER_4 = solve(MER_4)
###### inverse multiply Data transpose
        CER_4 = BER_4%*%QERt_4
###### Parameters of (1,1,2)
        dim(Q) = c(119,1)
        Qobs = Q
        PER_4 = CER_4%*%Qobs
        PER_4

###### Data transpose and multiply data matrix (2,1,2)
```

```
QER_5 = cbind(Q_1, Q_2, E_1, R_1, R_2)
dim(QER_5) = c(119,4)
QERd_5 = QER_5
QERt_5 = t(QER_5)
MER_5 = QERt_5%*%QERd_5
###### Inverse data
BER_5 = solve(MER_5)
###### inverse multiply Data transpose
CER_5 = BER_5%*%QERt_5
###### Parameters of (1,1,2)
dim(Q) = c(119,1)
Qobs = Q
PER_5 = CER_5%*%Qobs
PER_5                                    #### Selecting and Run
```

8. ###### Forecasting AR, ARMA AND ARMAX

```
###### (1,0,0)
QF1= P1%*%t(Q_1) +E
QF1
###### (2,0,0)
QF2 = (P2[1,] %*%t(Q_1)) +(P2[2,] %*%t(Q_2)) + E
QF2
###### (3,0,0)
QF3 = (P3[1,] %*%t(Q_1)) +(P3[2,] %*%t(Q_2)) +(P3[3,] %*%t(Q_3)) + E
QF3
###### (1,1,0)
QFE2 = (PE2[1,] %*%t(Q_1)) +(PE2[2,] %*%E_1) + E
QFE2
###### (2,1,0)
QFE3 = (PE3[1,] %*%t(Q_1)) +(PE3[2,] %*%t(Q_2)) +(PE3[3,] %*%E_1) + E
QFE3
###### (3,1,0)
```

```
QFE4 = (PE4[1,] %*%t(Q_1)) + (PE4[2,] %* %t(Q_2)) + (PE4[3,] % * %t(Q_3)) +
(PE4[4,] %*%E_1) + E
QFE4
###### (1,1,1)
QFER3 = (PER3[1,] %*%t(Q_1)) +(PER3[2,] %*%E_1) +(PER3[3,] %*%R_1) + E
QFER3
###### (2,1,1)
QFER4 = (PER4[1,] %*%t(Q_1)) +(PER4[2,] %*%Q_2) +(PER4[3,] %*%E_1)
+(PER4[4,] %*%R_1) + E
QFER4
###### (3,1,1)
QFER5 = (PER5[1,] %*%t(Q_1)) +(PER5[2,] %*%Q_2) +(PER5[3,] %*%Q_3)
+(PER5[4,] %*%E_1) +(PER5[5,] %*%R_1) + E
QFER5
###### (2,0,1)
QFR3 = (PR3[1,] %*%t(Q_1)) +(PR3[2,] %*%Q_2) +(PR3[3,] %*%R_1) + E
QFR3
###### (1,1,2)
QFER_4 = (PER_4[1,] %*%t(Q_1)) +(PER_4[2,] %*%E_1) +(PER_4[3,] %*%R_1)
+(PER_4[4,] %*%R_2) + E
QFER_4
###### (2,1,2)
QFER_5 = (PER_5[1,] %*%t(Q_1)) +(PER_5[2,] %*%Q_2) +(PER_5[3,] %*%E_1)
+(PER_5[4,] %*%R_1) +(PER_5[5,] %*%R_2) + E
QFER_5                    #### Selecting and Run
```

9.

```
###### Name the dataset forecasted and Plotting
dim(Q) = c(119,1)
Qobs = Q
colnames(Qobs) [1] <-"Observed_Flow"
Qf = Qobs
Qf<- data.frame (Qf)

dim(D) = c (119,1)
```

```
Dt = D
colnames(Dt) [1] <-"Date"
Date = Dt
Date <- data.frame(Date)

dim(QF1) = c(119,1)
QF1_D = QF1
colnames(QF1_D) [1] <-"p.1.0.0"
dim(QF2) =c (119,1)
QF2_D = QF2
colnames(QF2_D) [1] <-"p.2.0.0"

dim(QF3) = c (119,1)
QF3_D = QF3
colnames(QF3_D) [1] <-"p.3.0.0"

dim(QFE2) = c (119,1)
QFE2_D = QFE2
colnames(QFE2_D) [1] <-"p.1.1.0"

dim(QFE3) = c (119,1)
QFE3_D = QFE3
colnames(QFE3_D) [1] <-"p.2.1.0"

dim(QFE4) = c (119,1)
QFE4_D = QFE4
colnames(QFE4_D) [1] <-"p.3.1.0"

dim(QFER3) = c (119,1)
QFER3_D = QFER3
colnames(QFER3_D) [1] <-"p.1.1.1"

dim(QFER4) = c (119,1)
QFER4_D = QFER4
```

```
colnames(QFER4_D) [1] <-"p.2.1.1"

dim(QFER5) = c (119,1)
QFER5_D = QFER5
colnames(QFER5_D) [1] <-"p.3.1.1"

dim(QFR3) = c (119,1)
QFR3_D = QFR3
colnames(QFR3_D) [1] <-"p.3.0.1"

dim(QFER_4) = c (119,1)
QFER_4_D = QFER_4
colnames(QFER_4_D) [1] <-"p.1.1.2"

dim(QFER_5) = c (119,1)
QFER_5_D = QFER_5
colnames(QFER_5_D) [1] <-"p.2.1.2"

QF_2016sara = cbind (QF1_D, QF2_D, QF3_D, QFE2_D, QFE3_D, QFE4_D,
QFER3_D, QFER4_D, QFER5_D, QFR3_D, QFER_4_D, QFER_5_D)
QF_2016sara                    #### Selecting and Run

###### Plotting the forecasted dataset using ggplot2

DQ <- cbind (Date, Qf, QF_2016sara)
###### write.csv (DQ)
DQ <- data. frame (DQ)
DQ$Date <- as. Date (DQ$Date, "%m/%d/%Y")
ggplot (DQ, aes (Date, Observed_Flow, color = P_Forecasted)) + geom_line (colour =
"Blue", size = 1.2) + scale_x_date (labels = date_format("%b-%Y"), limits = c(as.Date
("2016-06-04"), as.Date("2016-09-30"))) + ylab ("Discharge in Cumecs") + xlab
("Date") + geom_line(data = DQ, aes(y = p.1.0.0, colour = "(1,0,0)"), size =0.8) +
geom_line(data = DQ, aes(y = p.2.0.0, colour = "(2,0,0)"), size =0.8) +
geom_line (data = DQ, aes (y = p.3.0.0, colour = "(3,0,0)"), size =0.8) +
```

```
geom_line (data = DQ, aes (y = p.1.1.0,   colour = "(1,1,0)"), size =0.8) +
geom_line (data = DQ, aes (y = p.2.1.0,   colour = "(2,1,0)"), size =0.8) +
geom_line (data = DQ, aes (y = p.3.1.0,   colour = "(3,1,0)"), size =0.8) +
geom_line (data = DQ, aes (y = p.1.1.1,   colour = "(1,1,1)"), size =0.8) +
geom_line (data = DQ, aes (y = p.2.1.1,   colour = "(2,1,1)"), size =0.8) +
geom_line (data = DQ, aes (y = p.3.1.1,   colour = "(3,1,1)"), size =0.8) +
geom_line (data = DQ, aes (y = p.1.1.2,   colour = "(1,1,2)"), size =0.8) +
geom_line (data = DQ, aes (y = p.2.1.2, colour = "(2,1,2)"), size =0.8) +
scale_y_continuous (limits = c (0,300)) + ggtitle ("Balganga river flow simulated") +
theme (plot.title = element_text (hjust = 0.5))          #### Selecting and Run
```

10. ###### Computing ACF, PACF residuals

```
acf (Qobs, lag.max = 20)
pacf (Qobs, lag.max = 20)
diffln_flow = diff (Qobs, 1)
acf (diffln_flow, lag.max = 20)
pacf (diffln_flow, lag.max = 20)          #### Selecting and Run
```

11. ###### Checking the Goodness of fit for the selected model and evaluation of the model performance

```
library(hydroGOF)
gof(QF1_D, Qobs)          #### Selecting and Run
```

DESCRIPTION VALIDATION FOR THE TIME SERIES ANALYSIS.

The dataset used in the program is Discharge and Rainfall. Date, discharge and rainfall are prepared in a spreadsheet and exported as CSV ("comma-separated value") file named 'SARA_ARMAX-2017.CSV'. For validation process, two spreadsheets of datasets were prepared and exported as CSV ("comma-separated value") file named 'SARA_ARMAX-2017.CSV' for calibration of parameters and 'SARA_ARMAX-2018_17val.CSV' for validation.

1. Setwd ("E:/WORK DIS/R/Sarasgaon")
2. Data = read.csv ('SARA_ARMAX-2017.CSV', TRUE, ",")

3. Data_18SA = read.csv ('SARA_ARMAX-2018_17val.CSV', TRUE, ",")

```
Attach(Data)
class(Data)
head(Data)

attach(Data_18SA)
class(Data_18SA)
head(Data_18SA)                                          #### Selecting all Run
```

4. ###### Installation of the R package in the library

```
library(ggplot2)      # Creat Elagant Data Visulisations Using Grammar of Graphics
library(MASS)         # Support Functions and Dataset for venables and Ripley's MASS
library(tseries)      # Time series Analysis and Computational Finance
library(forecast)     # Forecasting Functions for Time series and Linera Models
library(tidyverse)    # data manipulation and visualization
library(lubridate)    # easily work with dates and times
library(fpp2)         # Data for "Forecasting: Principles and Practice" (2nd Edition)
library(zoo)          # S3 Infrastructure for Regular and Irregular Time series
library(dplyr)        # A Grammar data Manipulation
library(scales)       # Scale Functions for Visualization
library(quantmod)     # Quantitative Financial Modelling Framework
library(readr)        # Read reactangular Text data
```

5. ###### Data preparation by lag using dataset of 2017.CSV for calibration of the parameters

```
Qf <- (Flow [4:125])
Q_1f <- (Flow [3:124])
Q_2f <- (Flow [2:123])
Q_3f <- (Flow [1:122])
Rf   <- (Rainfall [4:125])
R_1f <- (Rainfall [3:124])
R_2f <- (Rainfall [2:123])
R_3f <- (Rainfall [1:122])
```

```
Df <- (Date [4:125])

### using dataset of 2018.CSV for validation of the parameters
Q <- (Flow_18sar [4:125])
Q_1 <- (Flow_18sar [3:124])
Q_2 <- (Flow_18sar [2:123])
Q_3 <- (Flow_18sar [1:122])
R   <- (R_18sar [4:125])
R_1 <- (R_18sar [3:124])
R_2 <- (R_18sar [2:123])
R_3 <- (R_18sar [1:122])
D   <- (Date_18sar [4:125])
```

6.

```
###### Plotting the dataset using ggplot2 or normal graph plot
newdate <- as. Date (Data_18SA$Date_18sar, "%m/%d/%Y")
ggplot (Data_18SA, aes (x=newdate, y=Flow_18sar)) + geom_line (colour = "Blue") +
  scale_x_date (labels = date_format ("%b-%Y"), limits = c(as.Date ("2018-05-29"),
as.Date("2018-10-16")))+  ylab("Discharge in Cumecs") + xlab("Date")

plot (as.Date(Data_18SA$Date_18sar, "%m/%d/%Y"), Data_18SA$Flow_18sar, xlab
= "Dates", ylab = "Discharge in Cumecs", type = "l", col = "red", main = "Balganga
River flow simulated")
```

7.

```
###### Computing error
MAf <- rollmean (Flow, 7)     #Error
dim(MAf) = c(125,1)
Flf <- Flow [3:125]
dim(Flf) = c(123,1)
E1f   <- (Flf-MAf [2:124])
E_1f <- (E1f [1:122])
E_tf <- (E1f [2:123])
Ef   <-  E_tf

MA <- rollmean (Flow_18sar, 7)     #Error
dim(MA) = c(125,1)
```

```
Fl <- Flow_18sar [3:125]
dim(Fl) = c(123,1)
E1   <- (Fl-MA [2:124])
E_1 <- (E1 [1:122])
E_t<- (E1 [2:123])
E   <-  E_t
```

8. ###### Computing parameters (1,0,0), (2,0,0), (3,0,0), (1,1,0), (2,1,0), (3,1,0), (1,1,1), (2,1,1), (3,1,1), (1,1,1) by using matrix form.

```
###### Data transpose and multiply data matrix (1,0,0)
dim(Q_1f) = c(119,1)
Qd = Q_1f
Qt = t(Q_1f)
M1 = Qt%*%Qd
###### Inverse data
B1 = solve(M1)
###### inverse multiply Data transpose
C1 = B1%*%Qt
###### Parameter of (1,0,0)
dim(Q) = c(119,1)
Qobs = Q
P1= C1%*%Qobs
P1

###### Data transpose and multiply data matrix (2,0,0)
Q2 = cbind (Q_1f, Q_2f)
dim(Q2) = c(119,2)
Qd2 = Q2
Qt2 = t(Q2)
M2 = Qt2%*%Qd2
###### Inverse data
B2 = solve(M2)
###### inverse multiply Data transpose
C2 = B2%*%Qt2
```

```
###### Parameter of (2,0,0)
dim(Q) = c(119,1)
Qobs = Q
P2= C2%*%Qobs
P2
###### Data transpose and multiply data matrix (3,0,0)
Q3 = cbind(Q_1f, Q_2f, Q_3f)
dim(Q3) = c(119,3)
Qd3 = Q3
Qt3 = t(Q3)
M3 = Qt3%*%Qd3
###### Inverse data
B3 = solve(M3)
###### inverse multiply Data transpose
C3 = B3%*%Qt3
###### Parameter of (3,0,0)
dim(Q) = c(119,1)
Qobs = Q
P3= C3%*%Qobs
P3

###### Data transpose and multiply data matrix (1,1,0)
QE2 = cbind (Q_1f, E_1f)
dim(QE2) = c(119,2)
QEd2 = QE2
QEt2 = t(QE2)
ME2 = QEt2%*%QEd2
###### Inverse data
BE2 = solve(ME2)
###### inverse multiply Data transpose
CE2 = BE2%*%QEt2
###### Parameter of (1,1,0)
dim(Q) = c(119,1)
Qobs = Q
```

```
PE2= CE2%*%Qobs
PE2

###### Data transpose and multiply data matrix (2,1,0)
QE3 = cbind (Q_1f, Q_2f, E_1f)
dim(QE3) = c (119,3)
QEd3 = QE3
QEt3 = t(QE3)
ME3 = QEt3%*%QEd3
###### Inverse data
BE3 = solve(ME3)
###### inverse multiply Data transpose
CE3 = BE3%*%QEt3
###### Parameters of (2,1,0)
dim(Q) = c(119,1)
Qobs = Q
PE3= CE3%*%Qobs
PE3

###### Data transpose and multiply data matrix (3,1,0)
QE4 = cbind (Q_1f, Q_2f, Q_3f, E_1f)
dim(QE4) = c(119,4)
QEd4 = QE4
QEt4 = t(QE4)
ME4 = QEt4%*%QEd4
###### Inverse data
BE4 = solve(ME4)
###### inverse multiply Data transpose
CE4 = BE4%*%QEt4
###### Parameters of (3,1,0)
dim(Q) = c(119,1)
Qobs = Q
PE4= CE4%*%Qobs
PE4
```

```
###### Data transpose and multiply data matrix (1,1,1)
QER3 = cbind (Q_1f, E_1f, R_1f)
dim(QER3) = c(119,3)
QERd3 = QER3
QERt3 = t(QER3)
MER3 = QERt3%*%QERd3
###### Inverse data
BER3 = solve(MER3)
###### inverse multiply Data transpose
CER3 = BER3%*%QERt3
##### Parameters of (1,1,1)
dim(Q) = c(119,1)
Qobs = Q
PER3= CER3%*%Qobs
PER3

###### Data transpose and multiply data matrix (2,1,1)
QER4 = cbind (Q_1f, Q_2f, E_1f, R_1f )
dim(QER4) = c(119,4)
QERd4 = QER4
QERt4 = t(QER4)
MER4 = QERt4%*%QERd4
###### Inverse data
BER4 = solve(MER4)
###### inverse multiply Data transpose
CER4 = BER4%*%QERt4
###### Parameters of (2,1,1)
dim(Q) = c(119,1)
Qobs = Q
PER4= CER4%*%Qobs
PER4

###### Data transpose and multiply data matrix (3,1,1)
```

```
QER5 = cbind (Q_1f, Q_2f, Q_3f, E_1f, R_1f)
dim(QER5) = c(119,5)
QERd5 = QER5
QERt5 = t(QER5)
MER5 = QERt5%*%QERd5
###### Inverse data
BER5 = solve(MER5)
###### inverse multiply Data transpose
CER5 = BER5%*%QERt5
###### Parameters of (3,1,1)
dim(Q) = c(119,1)
Qobs = Q
PER5 = CER5%*%Qobs
PER5

###### Data transpose and multiply data matrix (2,0,1)
QR3 = cbind (Q_1f, Q_2f, R_1f)
dim(QR3) = c(119,3)
QRd3 = QR3
QRt3 = t(QR3)
MR3 = QRt3%*%QRd3
###### Inverse data
BR3 = solve(MR3)
###### inverse multiply Data transpose
CR3 = BR3%*%QRt3
###### Parameters of (2,0,1)
dim(Q) = c(119,1)
Qobs = Q
PR3 = CR3%*%Qobs
PR3

###### Data transpose and multiply data matrix (1,1,2)
QER_4 = cbind (Q_1f, E_1f, R_1f, R_2f)
dim(QER_4) = c(119,4)
```

```
QERd_4 = QER_4
QERt_4 = t(QER_4)
MER_4 = QERt_4%*%QERd_4
###### Inverse data
BER_4 = solve(MER_4)
###### inverse multiply Data transpose
CER_4 = BER_4%*%QERt_4
###### Parameters of (1,1,2)
dim(Q) = c(119,1)
Qobs = Q
PER_4 = CER_4%*%Qobs
PER_4

 ###### Data transpose and multiply data matrix (2,1,2)
QER_5 = cbind (Q_1f, Q_2f, E_1f, R_1f, R_2f)
dim(QER_5) = c(119,4)
QERd_5 = QER_5
QERt_5 = t(QER_5)
MER_5 = QERt_5%*%QERd_5
###### Inverse data
BER_5 = solve(MER_5)
###### inverse multiply Data transpose
CER_5 = BER_5%*%QERt_5
###### Parameters of (1,1,2)
dim(Q) = c(119,1)
Qobs = Q
PER_5 = CER_5%*%Qobs
PER_5
```

9.
```
###### Forecasting AR, ARMA AND ARMAX
###### (1,0,0)
QF1= P1%*%t(Q_1) +E
QF1
###### (2,0,0)
```

QF2 = (P2[1,] %*%t(Q_1)) +(P2[2,] %*%t(Q_2)) + E

QF2

(3,0,0)

QF3 = (P3[1,] %*%t(Q_1)) +(P3[2,] %*%t(Q_2)) +(P3[3,] %*%t(Q_3)) + E

QF3

(1,1,0)

QFE2 = (PE2[1,] %*%t(Q_1)) +(PE2[2,] %*%E_1) + E

QFE2

(2,1,0)

QFE3 = (PE3[1,] %*%t(Q_1)) +(PE3[2,] %*%t(Q_2)) +(PE3[3,] %*%E_1) + E

QFE3

(3,1,0)

QFE4 = (PE4[1,] %*%t(Q_1)) + (PE4[2,] %* %t(Q_2)) + (PE4[3,] % * %t(Q_3)) + (PE4[4,] %*%E_1) + E

QFE4

(1,1,1)

QFER3 = (PER3[1,] %*%t(Q_1)) +(PER3[2,] %*%E_1) +(PER3[3,] %*%R_1) + E

QFER3

(2,1,1)

QFER4 = (PER4[1,] %*%t(Q_1)) +(PER4[2,] %*%Q_2) +(PER4[3,] %*%E_1) +(PER4[4,] %*%R_1) + E

QFER4

(3,1,1)

QFER5 = (PER5[1,] %*%t(Q_1)) +(PER5[2,] %*%Q_2) +(PER5[3,] %*%Q_3) +(PER5[4,] %*%E_1) +(PER5[5,] %*%R_1) + E

QFER5

(2,0,1)

QFR3 = (PR3[1,] %*%t(Q_1)) +(PR3[2,] %*%Q_2) +(PR3[3,] %*%R_1) + E

QFR3

(1,1,2)

QFER_4 = (PER_4[1,] %*%t(Q_1)) +(PER_4[2,] %*%E_1) +(PER_4[3,] %*%R_1) +(PER_4[4,] %*%R_2) + E

QFER_4

(2,1,2)

```
QFER_5 = (PER_5[1,] %*%t(Q_1)) +(PER_5[2,] %*%Q_2) +(PER_5[3,] %*%E_1)
+(PER_5[4,] %*%R_1) +(PER_5[5,] %*%R_2) + E
QFER_5                                  #### Selecting and Run
```

10. ###### Name the dataset forecasted and Plotting

```
dim(Q) = c(119,1)
Qobs = Q
colnames(Qobs) [1] <-"Observed_Flow"
Qf = Qobs
Qf<- data.frame (Qf)

dim(D) = c (119,1)
Dt = D
colnames(Dt) [1] <-"Date"
Date = Dt
Date <- data.frame(Date)

dim(QF1) = c(119,1)
QF1_D = QF1
colnames(QF1_D) [1] <-"p.1.0.0"
dim(QF2) =c (119,1)
QF2_D = QF2
colnames(QF2_D) [1] <-"p.2.0.0"

dim(QF3) = c (119,1)
QF3_D = QF3
colnames(QF3_D) [1] <-"p.3.0.0"

dim(QFE2) = c (119,1)
QFE2_D = QFE2
colnames(QFE2_D) [1] <-"p.1.1.0"

dim(QFE3) = c (119,1)
QFE3_D = QFE3
```

```
colnames(QFE3_D) [1] <-"p.2.1.0"

dim(QFE4) = c (119,1)
QFE4_D = QFE4
colnames(QFE4_D) [1] <-"p.3.1.0"

dim(QFER3) = c (119,1)
QFER3_D = QFER3
colnames(QFER3_D) [1] <-"p.1.1.1"

dim(QFER4) = c (119,1)
QFER4_D = QFER4
colnames(QFER4_D) [1] <-"p.2.1.1"

dim(QFER5) = c (119,1)
QFER5_D = QFER5
colnames(QFER5_D) [1] <-"p.3.1.1"

dim(QFR3) = c (119,1)
QFR3_D = QFR3
colnames(QFR3_D) [1] <-"p.3.0.1"

dim(QFER_4) = c (119,1)
QFER_4_D = QFER_4
colnames(QFER_4_D) [1] <-"p.1.1.2"

dim(QFER_5) = c (119,1)
QFER_5_D = QFER_5
colnames(QFER_5_D) [1] <-"p.2.1.2"

QF_2016sara = cbind (QF1_D, QF2_D, QF3_D, QFE2_D, QFE3_D, QFE4_D,
QFER3_D, QFER4_D, QFER5_D, QFR3_D, QFER_4_D, QFER_5_D)
QF_2016sara                          #### Selecting and Run
```

Plotting the forecasted dataset using ggplot2

```
DQ <- cbind (Dte, Q_f, QF_2018Sara)
###### write.csv (DQ)
DQ <- data. frame (DQ)
DQ$Dte <- as. Date (DQ$Dte, "%m/%d/%Y")
ggplot (DQ, aes (Dte, Observed_Flow, color = P_Forecasted)) + geom_line (colour = "Blue", size = 1.2) + scale_x_date (labels = date_format("%b-%Y"), limits = c(as.Date("2018-06-01"), as.Date("2018-09-30"))) + ylab("Discharge in Cumecs") + xlab("Date") + geom_line(data = DQ, aes(y = p.1.0.0, colour = "(1,0,0)"), size=0.8) +
 geom_line (data = DQ, aes (y = p.2.0.0, colour = "(2,0,0)"), size=0.8) +
geom_line (data = DQ, aes (y = p.3.0.0, colour = "(3,0,0)"), size =0.8) +
geom_line (data = DQ, aes (y = p.1.1.0, colour = "(1,1,0)"), size =0.8) +
geom_line (data = DQ, aes (y = p.2.1.0, colour = "(2,1,0)"), size =0.8) +
geom_line (data = DQ, aes (y = p.3.1.0, colour = "(3,1,0)"), size =0.8) +
geom_line (data = DQ, aes (y = p.1.1.1, colour = "(1,1,1)"), size =0.8) +
geom_line (data = DQ, aes (y = p.2.1.1, colour = "(2,1,1)"), size =0.8) +
geom_line (data = DQ, aes (y = p.3.1.1, colour = "(3,1,1)"), size =0.8) +
geom_line (data = DQ, aes (y = p.1.1.2, colour = "(1,1,2)"), size =0.8) +
geom_line (data = DQ, aes (y = p.2.1.2, colour = "(2,1,2)"), size =0.8) + scale_y_continuous (limits = c (0,300)) + ggtitle ("Balganga river flow simulated”) + theme (plot.title = element_text (hjust = 0.5))
```

Selecting and Run

11. ###### Computing ACF, PACF residuals

```
acf (Qobs, lag.max = 20)
pacf (Qobs, lag.max = 20)
diffln_flow = diff (Qobs, 1)
acf (diffln_flow, lag.max = 20)
pacf (diffln_flow, lag.max = 20)
```

Selecting and Run

12. ###### Checking the Goodness of fit for the selected model and evaluation of the model performance

```
library(hydroGOF)
gof(QF1_D, Qobs)
```

Selecting and Run