FLOOD FREQUENCY ANALYSIS OF PARTIAL DURATION SERIES USING SOFT COMPUTING TECHNIQUES FOR MAHANADI RIVER BASIN IN INDIA

A THESIS SUBMITTED

FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

CIVIL ENGINEERING

BY

NIBEDITA GURU
(ROLL NO. 512CE102)

UNDER THE SUPERVISION OF
Prof. RAMAKAR JHA

NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA-769008, INDIA

JANUARY-2016

National Institute of Technology Rourkela
Rourkela-769008

CERTIFICATE

This is to certify that the thesis entitled "**Flood Frequency Analysis of Partial Duration Series Using Soft Computing Techniques for Mahanadi River Basin in India"** being submitted by **Ms. Nibedita Guru** for the award of the degree of Doctor of Philosophy (Civil Engineering) of NIT Rourkela, is a record of bonafide research work carried out by her under my supervision and guidance.

To the best of our knowledge, the matter embodied in the thesis has not been submitted to any other university or institution for the award of any degree or diploma.

Date:

Place: Rourkela

Prof. Dr. Ramakar Jha
Department of Civil Engineering

Dedicated to

My Parents

ACKNOWLEDGEMENT

First and foremost I convey my thanks to National Institute of Technology, Rourkela for providing me the opportunity to study for the degree of Doctor of Philosophy in Civil Engineering.

I feel glad to offer my sincere gratitude and respect to my supervisor Prof. Ramakar Jha, Water Resources Engineering group of Civil Engineering Department for his precious advice, and guidance from the preliminary stage of this research and providing me extraordinary experiences throughout the work. I truly appreciate his esteemed guidance and encouragement from beginning to the end of the thesis.

I wish to express my sincere thanks to Prof. S.K. Sahu, Chairman of doctoral committee and the doctoral committee members Prof. K. K. Khatua, Prof. H.B. Sahu and Prof. S. C. Mohanty, for their valuable suggestions and instructions in my work.

I wish to express my sincere gratitude to Dr. S. K. Sarangi, Director, NIT, Rourkela and Prof. S.K. Sahu, Head of the Department, Civil Engineering, for giving me the opportunities, the academic support and facilities to carry out my research work.

I would like to express my gratitude to all the faculty and staff members of the Department of Civil Engineering, National Institute of Technology, Rourkela for their continuous support and moral inspiration.

I am also obliged to Sanjita Jaipuria, Janhabi Meher, Raysingh Meena, Brijesh Kumar, Janaki Ballav Swain, Mrunmayee Manjari Sahoo and Bandita Nayak for their support and co-operation that is difficult to express in words. The time spent with them will remain in my memory for years to come.

Last but not the least, I am especially indebted to my parents and sisters for their unwavering support, love and consistent source of motivation.

Nibedita Guru

ABSTRACT

In flood frequency analysis, the modeling based on Annual Maximum Flood (AMF) series remains the most popular approach. An alternative approach based on the "partial duration series (PDS) or peaks over threshold (POT)" has been considered in recent years, which captures more information about extreme events by fixing appropriate threshold values. The PDS approach has lot of advantages, (i) it consist more peak events by selecting the appropriate threshold hence to capture more information regarding the flood phenomena. (ii) it analyses both, the time of arrival and the magnitude of peaks, (iii) it provides extra flexibility in the demonstration of floods and a complete explanation of the flood generating process. However, the PDS approach remains underused and unpopular due to the nonexistence of general framework regarding different approaches.

The first objective of the present research work is to develop a framework in the above question on selection of an appropriate threshold value using different concepts and, to verify the independency and stationarity criteria of the extreme events for the modeling of the PDS in the Mahanadi river system, India. For the analysis, daily discharge data from 22 stations with record length varying between 10 and 41 years have been used with the assumption that the whole basin is homogeneous in nature. The results confirmed that the Generalized Pareto (GP) best described the PDS in the study area and also, show that the best PDS/GP performance is found in almost all the value of λ (2, 2.5 and 3).

In the second phase, the analysis is done to carry out the regional flood frequency analysis in the Mahanadi basin and to apply the developed model to the respective homogeneous region. Regionalization is the best viable way of improving flood quantile estimation. In the regional flood frequency analysis, selection of basin characteristics, morphology, land use and hydrology have significant role in finding the homogeneous regions. In this work the Mahanadi basin is divided into homogeneous regions by using fifteen effective variables initially. However, it has been observed that the whole basin is not hydro meteorologically homogeneous. Therefore, Factor analysis has been introduced in finding suitable number of variables, and nine variables are found suitable for analysis. Hierarchical (HC) and K-Means Clustering (KM) techniques are used for finding out the possible number of clusters. Here, again the Generalized Pareto (GP) distribution best described the PDS in the study area. To test the homogeneity and to identify the best-fit frequency distribution, regional L-moment algorithm is used. A unique regional flood frequency curve is developed which can estimate the flood quantiles in ungauged catchments and an index flood is also specified concerning the catchment characteristics by using the multiple linear regression approach.

In the third and fourth phase, to demonstrate the rainfall, corresponding peak flows obtained using PDS and resultant flood inundated area, many models (ANN, ANFIS, HEC-GeoRAS and HEC-RAS models) developed in recent past are used and have been tested for their applicability in Lower Mahanadi river basin. It has been observed that the floods and inundation due to these peak rainfall and discharge typically depend on various parameters including time of concentration, basin slope, river morphological characteristics, rainfall, soil moisture, groundwater, land use, and river discharge during monsoon. For this reason, a different combination of parameters have been used to obtain the best model. The ANN and ANFIS models have been used to estimate runoff occurred due to corresponding discharge and rainfall whereas HEC-RAS has been used to estimated flood inundation due to different magnitude of peak flows and corresponding water level. Moreover, a refined coupled model has been developed to estimate high flow above a threshold value and the resultant inundated areas using remote sensing and GIS. Remote sensing images are effective tools to determine the spatio-temporal flood extents. The goodness of fit values and correlation statistics has been tested using different error criteria. The results obtained using coupling mechanism is found to be very useful for estimating both, the high flow above a threshold value and the corresponding inundation due to floods.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AMS- Annual Maximum Series

PDS- Partial Duration Series

λ- Average Number of Peaks per Year

HFL- Highest Flood Level

GP- Generalized Pareto Distribution

t- Threshold Value

DI- Dispersion Measure

μ- Location Parameter

σ- Scale Parameter

ξ- Shape Parameter

LMRD- L-moment Ratio Diagram

RRMSE- Relative Root Mean Square Error

RBIAS- Relative Bias

FCM- Fuzzy C Mean

NP- Nondeterministic Polynomial Time

QQ Plot- Quantile-Quantile Plot

HC-Hierarchical Cluster

E- Elevation

SL- Average Slope

AMP- Annual Mean Precipitation

P- Perimeter

CR- Compactness ratio

FF- Form factor

SF- Shape Factor

D_i- Discordance Measure

PCs- Principal Components

RNG- Random Number Generation

ANN- Artificial neural Network

HN- Hidden Neurons

LM- Levenberg-Marquardt

GDX- Gradient Descent Algorithm

CGF- Conjugate Descent Algorithm

ANFIS- Artificial Neuro Fuzzy Inference System

DEM- Digital elevation Model

TIN- Triangular Irregular Network

DTM- Digital Terrain Model

NDWI- Normalised Difference Water Index

SWIR- Short- Wave Infrared Radiation

MODIS- Moderate Resolution Imaging Spectroradiometer

H- Energy Head

h_f – Friction Head Loss

K- Conveyance

S_f- Average Friction Slope

R- Hydraulic Radius

n- Manning's Roughness

h_o- Contraction or Expansion Head Loss

EGL- Energy Grade Line

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND

Floods are one of the most frequently occurring natural disasters in Lower Mahanadi river basin in India. They occurs mainly due to the inadequate water holding capacity of the river, low retention capacity of the flood plain during monsoon, heavy rainfall, sudden release of water from Hirakud dam and varying time of concentration. Such complexity inundates the densely populated area in lower Mahanadi river basin every year, causing immense loss to the life and property. After commissioning of Hirakud dam during 1958, flood miseries have reduced to some extent. However, due to the sudden release of water during monsoon causes flooding. Most recent floods in lower Mahanadi basin occurred during the year 2008, 2011 and 2013.

To prevent the losses in the lower Mahanadi basin and to assess the periodic flood inundation, flood frequency analysis for different return periods at regional and local scale; and rainfall runoff modeling is essential. Accurate prediction of the flood inundation area is essentially required for developing and quantifying flood insurance rates. Many researchers have done studies using different hydro-climatic variables for Mahanadi river basin (Rao, 1993, 1995; Gosain et al., 2006; Raje and Mujumdar, 2009; Asokan and Dutta, 2008; Ghosh et al., 2010). In addition, some studies are done for trends analysis of flood series (Birsan et al., 2005; Kumar et al., 2009; Zhang et al., 2010). Gosain et al., (2006) predicted that the Mahanadi river basin would be the worst affected basin due to the change of climate which may cause the occurrence of high-intensity floods in the basin. They stated that the peak flood at the middle reach of the basin would exceed from 20,000 m^3/s in controlled scenario (1981–2000) to 37,000 m^3/s in the future Green House Gases (GHG) scenario (2041–2060). Ghosh et al., (2010) studied climate models in Mahanadi River and observed a decreasing flow trend at Hirakud dam under a future climate scenario.

Although, some studies have been carried out in the Mahanadi river basin using hydro-climatic variables, a detailed analysis to assess the flood frequency for different return periods and their impact on flood inundation is lacking. In the present work, a detailed study is done using different input variables including peak discharges of different sites, regulated release of flow from the Hirakud dam, corresponding peak rainfall and land use to study the flood frequency of different return periods at local and regional levels using partial duration series (PDS), flood forecasting using Artificial Neural Networks (ANN) and Artificial Neuro Fuzzy Inference System (ANFIS), and assessing corresponding flood inundation in lower Mahanadi river

system using Hydrologic Engineering Center- River Analysis System (HEC-RAS) models and Geographical Information System (GIS) platform.

1.2 At-Site Flood Frequency Analysis Using Partial Duration Series

Flood frequency analysis is used to predict flood magnitudes having different return periods at various sites in a river. The analysis consists of both annual peak flow discharge data and the partial duration series flood discharge data above a threshold value to compute statistical information such as mean values, standard deviations, skewness, and recurrence intervals. These statistical data are then used to make frequency distributions for various discharges as a function of their recurrence interval or exceedance probability (Hosking and Wallis, 1997). Developing relationship among flood frequencies and corresponding return periods are essential for designing and safeguarding many hydraulic structures such as dams, barrages, check-dams, culverts and urban drainage systems (Stedinger et al., 1992; Meng et al., 2007). Appropriate flood frequency analysis would certainly provide a mechanism to control, manage and predict floods and their impact in gauged as well as ungauged regions of a basin.
In flood frequency analysis, different opinions exist concerning reasonable merits of sampling a random sequence of extreme values either as a partial duration series (PDS) or annual maximum series (AMS). Figure 1.1 demonstrates the difference between PDS and AMS for estimating flood frequency.

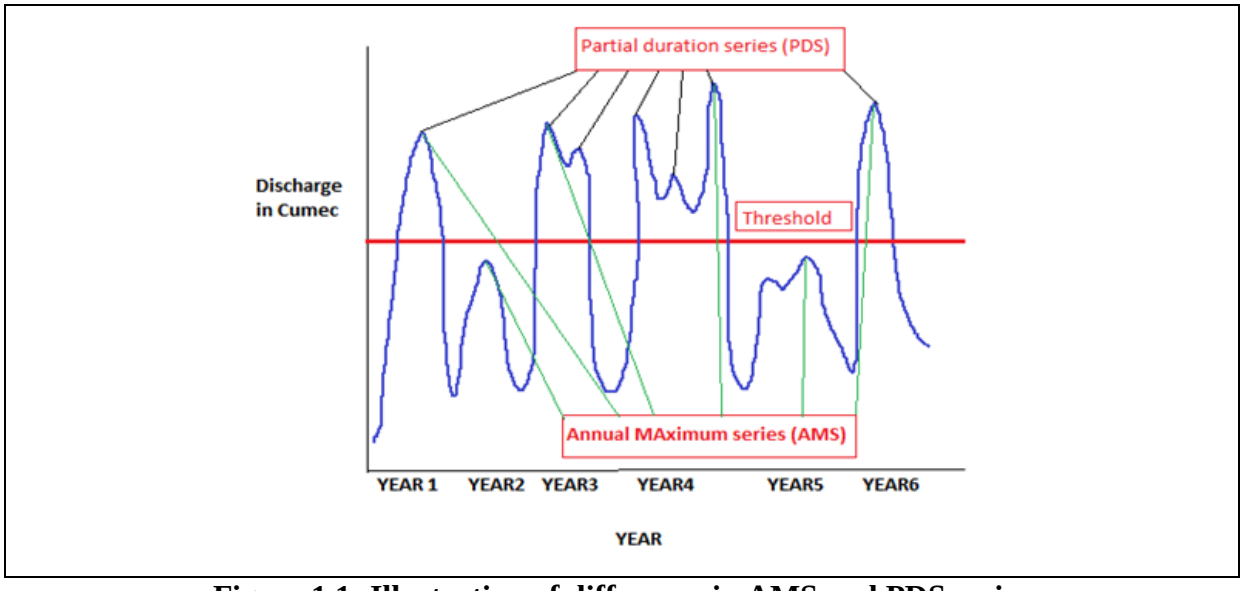


Figure 1.1: Illustration of difference in AMS and PDS series

A minimum of 30-40 years of records is needed for flood frequency analysis. The selection of a probability distribution is of fundamental importance in flood frequency analysis, as a wrong

choice could lead to significant error and bias in design flood estimates. Particularly at higher return periods, the flood may result in either under estimation or over-estimation, which may have serious implications in practice. There have been numerous studies in the past on the comparison of various probability distributions for on-site flood frequency analysis. A considerable realization of the late 20th century in statistical estimation was the method of L-moments. They are analogous to conventional moments but are estimated as linear combinations of order statistics. Nowadays, this method is widely used for estimating various hydrometeorological variables and for estimating distribution parameters.

1.3 Regional Flood Frequency Analysis Using Partial Duration Series

Basin regionalization typically refers to a grouping of basins or sub-basins into homogeneous regions, which contain sites of similar flood producing characteristics and is an important focus as most structures are constructed in the flood prone areas where recorded flood data are either missing or inadequate. Regional flood frequency methods include the (i) Index Flood method (Dalrymple, 1960; Hosking and Wallis, 1988, 1997; Stedinger and Lu, 1995; Fill and Stedinger, 1998; De Michele and Rosso, 2001), in which the flood regime comprises of the magnitude, timing, duration, frequency, and inter-annual predictability of flood events considered similar to allow the spatial transfer of information from gauged sites to ungauged sites and (ii) Regional regression procedures, such as weighted and generalized least squares regression (Tasker and Stedinger 1989; Tasker et al., 1996; Madsen and Rosberg 1997; Eng et al., 2005, 2007; Griffis and Stedinger 2007a), in which the knowledge of the physical properties and mechanisms producing flood flows remain limited. In fact, the regression modeling approach is to employ a log-log or log-linear relationship between flood statistics and catchment characteristics, and in most cases the drainage area is the only descriptive variable used. However, the complex relationships between catchment characteristics are avoided due to the inadequate understanding of the hydrology of a river basin. Thus, by improving the knowledge regarding the catchment characteristics, various statistical analysis and flood quantiles estimation at ungauged sites can be developed. It may also be possible to improve the quantile estimation by using the remotely sensed data in a limited data situation.

Many researchers have performed regional frequency analysis of annual maximum flows to predict extreme flow for the future, but there are no significant applications of regional frequency analysis using partial duration flow series as they describes extreme events in a better ways in comparison to the annual maximum series model (Buishand, 1989; Norbiato et al., 2007).

1.4 Flood Forecasting

In the recent years, many models have been developed based on physical and theoretical approaches for river flow forecasting (Aqil et al., 2007; Mukerjee et al., 2009). However, it is not easy to predict the flow of a river system using the traditional flood routing methods, if the flow control structures are present. Hydrologists have commonly used the upstream discharge and water level to predict downstream discharges. In recent years, soft computing techniques such as artificial neural networks (ANN) are increasingly used for forecasting floods (Karunanithi et al., 1994; Thirumalaiah and Deo, 1998; Dawson and Wilby, 1998; Zealand et al., 1999; Chang et al., 2002; Sivakumar et al., 2002; Lekkas et al., 2001). Although ANN techniques are proven to be effective, they have been strengthened further using neuro-fuzzy inference systems. The neuro-fuzzy method is superior because it is able to acquire the information of both ANN and fuzzy logic in a single framework.

1.5 Flood Inundation

Flood inundation modeling is the process of defining the area filled with water during the flooding periods. When this flooded area is represented on a map it is referred as the flood inundation mapping. Flood inundation modeling comprises both hydrological and hydraulic modeling (Anderson, 2000; Robayo et al., 2004; Knebl et al., 2005). Hydrological modeling estimates the peak flows from flood events whereas the hydraulic modeling estimates the water surface elevations and the resulted flood inundated areas using digital terrain model. Generally, in low-frequency events, the data are not be available sometimes and estimation of flood inundated areas for low return period flood events is seldom possible. To prevent the losses, a reliable information regarding the risk associated with the flooding is provided to the public, emergency managers and city planners. As flood occurrences and their serious consequences are common in various parts of the world, it has raised public, political and scientific awareness for proper flood control and management (Becker et al., 2003). Using non-structural techniques, valuation and management of flood inundated area for different magnitudes of floods are very crucial. Various hydrologic models have been developed in the past to simulate flood inundation in the basin areas (Iwasa and Inoue, 1982; Samules, 1985; Gee et al., 1990). These models consider overland and river flows. Only a few models such as HEC-GeoHMS, HEC-GeoRAS, MIKE BASIN, MIKE-11, MIKE-FLOOD are available to simulate flood inundation in a river basin for real flood events considering all the spatial heterogeneity of physical characteristics of topography.

1.6 Significance of the Study

The Mahanadi river basin covering major portions of Chhattisgarh and Orissa has been repetitively facing adverse hydro- meteorological conditions such as floods, droughts and cyclones in the recent times. The river has often been referred to as the 'Sorrow of Orissa'. The inhabited inner basin Chhattisgarh plain is suffering from frequent droughts whereas the fertile deltaic area has been subjected to repeated floods despite the operation of dams and barrages to control them. The hydrology has changed considerably due to the increased anthropogenic activities, producing disasters. The frequent occurrence of these events indicate a shift in the hydrological response of the basin because of which the upstream of Hirakud Dam is unable to retain sufficient moisture resulting in drought, and the downstream river is unable to handle large streamflows resulting in floods. The reason for such changes in hydrological regime could be attributed to the long-term climate change and landuse/ landcover changes in the region. The landuse/ landcover change impact assessment on hydrology more specifically streamflows can be best handled through simulation of the hydrological conditions that shall prevail under the projected weather conditions in an area. Such a treatment is essential because of the fact that the hydrological response is an extremely complex process governed by a large number of variables such as terrain, land use, soil characteristics and the state of the moisture in the soil.

Keeping this in view, a comprehensive study has been done in Mahanadi river system with the following objectives.

1.7 OBJECTIVE OF THE STUDY

1. To evolve a framework on appropriate selection of threshold value using different concepts for the on-site flood frequency modeling using Partial Duration Series in the Mahanadi river system, Odisha, India considering the daily discharge data from 22 stations with the record length varying between 10 and 41 years.

2. To develop a regional flood frequency analysis using partial duration series for Mahanadi river basin, India using basin characteristics, morphological, and hydrological data sets.

3. To examine the prediction accuracy of ANN and ANFIS based approaches for predicting peak discharge in the lower catchment of Mahanadi river basin prone to floods and flood inundation.

4. To construct a HEC-RAS model for generating flood inundation and flood risk map for the delta region of Mahanadi basin in relation to partial duration series of flood events with the support of remote sensing and GIS techniques.

1.8 OUTLINE OF THE THESIS

In view of the above objectives the present work has been divided into different Chapters as indicated below:

Chapter 1 describes the brief introduction of the on-site flood frequency analysis of partial duration series, regional flood frequency analysis, flood forecasting, flood inundation, objectives of the research, significance and outline of the chapters.

Chapter 2 provides a comprehensive literature review on the use of both on-site and regional flood frequency analysis using partial duration series, flood forecasting techniques and flood inundation modeling.

Chapter 3 describes various salient features of Mahanadi river basin and various data used for the above analysis.

Chapter 4 presents a framework for the selection of threshold in partial duration series modeling using different concepts and after selection of threshold value, modeling of PDS with different criteria of Mahanadi river basin is carried out.

Chapter 5 examines the regional flood frequency analysis of Mahanadi river basin using partial duration series to find out the homogeneous region by applying different techniques. Then it moves on to develop regional flood frequency curve for estimating the flood quantiles with different return periods for the homogeneous region.

Chapter 6 gives the application of soft computing techniques for river flow prediction in the lower catchment of Mahanadi river basin using partial duration series.

Chapter 7 includes the flood inundation and 1-d hydrodynamic modeling using remote sensing and GIS technique in the delta region of Mahanadi river basin, India.

Chapter 8 provides the detailed summary, conclusions, and future scope of the work.

CHAPTER 2
REVIEW OF LITERATURE

2.1 BACKGROUND

The flood frequency analysis for any river location is mostly done to estimate the flood for the desired recurrence interval, assuming that the sample data, which is a true representative of the population, follows a theoretical frequency distribution. Over the last century, various techniques have been developed to estimate the design flood using flood frequency analysis techniques. In this chapter, a critical appraisal of literature survey has been done on various aspects of flood frequency analysis using partial duration series, regional flood frequency analysis using partial duration series, flood forecasting using Artificial Neural Network and Neuro-Fuzzy Inference System and flood inundation modeling to provide a closer look at the growth, development, gaps, and application of these models.

2.2 On-Site Flood Frequency Analysis using Partial Duration Series

2.2.1 Historical Development in Flood Frequency Analysis Techniques

People have preferred to live along the river banks since the beginning of the human civilization, and therefore flood studies have been done by stakeholders for centuries to protect their livelihoods.

In the year 1868, O'Connell performed one of the earliest studies on regional analyses of stream flows with simple empirical formula that attempted to connect discharge to drainage area. The approach was very simple, and the proposed formula was

$$Q_{max} = CA^{0.5} \quad (2.1)$$

Where Q_{max} = maximum discharge; A = drainage area; and C = coefficient related to the region.

The application of probability theory in flood estimation procedures was introduced by Fuller (1914) for catchments in the U.S. in this study, the average of the maximum floods ( ) was related to the drainage area with an exponent of 0.8

$$\bar{Q} = CA^{0.8} \quad (2.2)$$

Further, an attempt to relate the discharge of a specified return period to drainage area was done. In

addition to equation (2.2), Fuller (1914) presented other formula relating annual maximum daily flows to drainage area for a given return period, T, as

$$Q_T = CA^{0.8}[1 + 0.8\log_{10}(T)](1 + 2A^{-0.3}) \quad (2.3)$$

Where Q_T = peak discharge in cubic feet per second with a return period of T years; C = coefficient related to the region; and A = drainage area in square miles.

Fuller (1914) also used plotting positions for analyzing flood distributions and suggested plotting at the median, which later on became popular as Hazen plotting position analysis. Many similar empirical relationships have been developed to relate discharge to drainage-basin characteristics.

Hazen (1914) recommended the use of logarithmic probability paper on which the log-normal (LN) distribution would plot as a straight line. It was found that the observed annual maximum flow series would plot as a straight line on the logarithmic probability paper than on a normal probability paper, thereby indicating that the LN distribution provided a better fit to the data.

Hazen (1921) revised an earlier work and found some datasets plotted as curved lines in log-normal distribution. Hence it was suggested to use a three-parameter distribution including skewness and plot it on a logarithmic probability paper.

Foster (1924) introduced the Pearson type III (P3) distribution for describing the flood data.

Kinnison (1930) raised the interest on flood hydrology based on the New England flood of 1927 by United States Geological Survey (USGS). The interest generated by that flood and the funds provided by the USGS for studies, resulted in the classic work by Jarvis (1926) on flood-frequency analysis and its companion work by Hoyt (1936) on rainfall-runoff analysis.

Gumbel (1941) brought the basis of analysis to a new level by applying extreme value theory and introduced the Extreme Value Type I distribution (EV1) to flood frequency analysis.

Benson (1968) compared the most commonly used methods of flood frequency analysis. Based on probability plots, it was recommended that the flood of given frequency is estimated by fitting the Log Pearson Type 3 (LP3) distribution to the series of annual maximum floods and that all U.S Government agencies adopt this as their base method in order to achieve a uniform procedure for estimating design floods.

The method of moments can be applied in two ways to estimate parameters of the Log Pearson type III distribution. Bobee and Robitaille (1975) proposed that the method of moments be applied directly to the observed data (i.e., direct method). This method is named as the Method of Bobee (MOB). First three moments about zero are used to estimate parameters in MOB.

Chow et al., (1988), related the magnitude of extreme events with their frequency of occurrence through the use of probability distributions. For ungauged catchments, the regional flood frequency analysis approach was found to be the only method to estimate the flood discharge for the desired recurrence interval.

Vogel and Wilson (1996) studied extensively the data of 1455 sites for the selection of probability distribution of annual maximum, mean and minimum streamflows in the US with the help of L-moment diagrams. Their study revealed that annual minimum streamflows in the US are best approximated by the Pearson type 3 distribution, whereas Pearson type 3, Log-Normal (3-parameter) and Log Pearson type 3 distributions provide a better fit for annual average streamflows.

Zafirakou-Koulouris et al. (1998) have mentioned that like ordinary product moments, L-moments summarized the characteristics or shapes of theoretical probability distributions and observed samples.

2.2.2 Partial Duration Series in Flood Frequency Analysis

Cunnane (1973) described a method for comparing the statistical efficiency of the T-year flood Q (T) estimation. They found that the estimated flood using partial duration series contains at least 1.65 N items, where N is the number of years of record having smaller sampling variance as compared to the annual maxima series for the same range of return periods.

Cunnane (1979) examined the validity of the Poisson distribution, using data from 26 gaging stations on 20 catchments in Great Britain. It was observed that when all the data are considered jointly, the Poisson assumption has to be rejected although it is acceptable in some cases. It was also suggested that if dependence exists in the partial duration series it should be considered for in the point process.

Ashkar et al. (1983) mentioned the truncation level above which streamflow is considered as flood flow play an important role in the partial duration series approach. They base their numerical investigation on some commonly used partial flood series models to show that once

the time-dependent Poisson process, used in modeling flood frequency is found applicable to a certain truncation level, then it should remain so with any higher truncation level. They also pointed out that this same property holds true for the exponential distribution widely used in the study of flood magnitude.

Takeuchi (1984) reviewed the Langbein's formula derived in 1949, which relates the hydrological recurrence intervals calculated from an annual maximum series and from a partial duration series, and presented an alternative derivation procedure with proper validation.

Cunnane (1985) discussed theoretical arguments on the empirical criteria for selection of distribution and previously used methods of discrimination between candidate distributions for modeling flood series which may give markedly different magnitude-return period (Q-T) relationships, especially at high T values. The impact of method of parameter estimation, treatment of outliers, and the inclusion of large historical flood values, data transformations and inventive composition of the flood population on choice of distribution were also considered.

Hosking and Wallis (1987) discussed generalized Pareto distribution which is a two-parameter distribution that contains uniform, exponential, and Pareto distributions as special cases and has various applications in a number of fields. They found that the parameters of generalized Pareto distribution derived from the method of moments or the method of probability weighted moments were more reliable than the maximum likelihood estimation method.

Bobee et al., (1993) reviewed the commonly used procedures for flood frequency estimation, pointed out some of the reasons for the present state of confusion concerning the advantages and disadvantages of the various methods, and proposed the broad lines of a possible comparison strategy. They recommend that the results of such comparisons be discussed in an international forum of experts, with the purpose of attaining a more coherent and broadly accepted strategy for estimating floods.

Wilks, (1993) investigated the performance of three-parameter probability distributions for representing annual extreme and partial duration precipitation data at stations in the northeastern and southern United States. They found that the beta-κ distribution best describes the extreme right tail of annual extreme series, and the beta distribution was the best for the partial duration data.

Pearson et al., (1998) presented a map identifying regional tendencies toward EV2 rainfall

distributions which can be used to supplement Tomlinson's method for estimating annual maximum storm rainfall frequencies in New Zealand.

Langa et al., (1999) reviewed that flood frequency analysis using annual maximum flood (AMF) was the most popular. They also used an alternative approach based on the "peak over threshold" (POT) approach. They proposed different tests for selection of threshold values and verified their independency and stationarity, and also presented an application.

Onoz et al., (2001) analyzed partial duration series (PDS) with POT as an alternative to annual maxima series in flood frequency analysis. They obtained various expressions for the estimation of the T-year flood and its sampling variance when binomial (or negative binomial) model was combined with the exponential distribution of peak magnitudes. They found that the results were almost identical to those obtained using the Poisson model, for which much simpler expressions were available.

Claps et al., (2003) proposed a filtered peaks over threshold (FPOT) procedure as an alternative to the PDS approach for determining the average annual number of flood events λ and applied to 33 time series data of daily runoff from rivers of northwestern Italy. The revised procedure demonstrate that there was no need for specific limitations on the magnitude of λ to preserve the fundamental hypotheses of the marked point process built in the PDS procedure.

Rosbjerg et al., (2004) summarized the important extensions of the PDS/POT method since the mid-1990s. The PDS/Generalised Pareto (GP) model was shown to be competitive with the AMS/Generalised Extreme Value (GEV) model and highly efficient for regionalization. They developed new procedures for testing the basic assumptions, introduced Generalised maximum likelihood and developed filtering methods for selection of independent threshold exceedances. They demonstrated the strengths of Bayesian methods in PDS analysis. The method was attractive for analysis of extreme hydrological events.

Begueria (2005) found that the partial duration series modeling was a robust tool for modeling of hydrologic extremes, but it remained underused due to several technical problems. The most important difficulty was the choice of the threshold value which affects the basic assumptions including arrival times and exceedance magnitudes. They considered the changes in parameter and quantile estimation as a function of the threshold value. They used simulated and real data to test the consistency of the model, and proposed a new modeling procedure based on increasing threshold censoring to overcome these problems.

Meng et al., (2007) presented an ERM-POT (Exponential Regression Model and the Peaks-Over-Threshold) method to measure the operational risk which had become increasingly important topics for Chinese Commercial Banks in recent years. Considering the huge operational losses, Extreme value theory (EVT) has been recognized as a useful tool for analyzing such data. They found that the ERM-POT method can lead to bias-corrected estimators and techniques for optimal threshold selections, and also the experimental results showed that the method is reasonable.

Deidda et al., (2009) analyzed the daily rainfall time series highlighted the presence of records with anomalous rounding (1 and 5 mm) while the standard resolution should be 0.1 or 0.2 mm. Assuming that the generalized Pareto distribution (GPD) can reliably represent the distribution of daily rainfall depths, they investigated how such discretizations can affect the inference process. The performance of several GPD estimators are compared using the Monte Carlo approach. Synthetic samples were drawn by GPDs with shape and scale parameters in the range of values estimated for the daily rainfall time series. They found how the relative efficiency of estimators could be very different for continuous or rounded-off samples.

Zvi (2009) proposed a procedure for basing intensity–duration–frequency (IDF) curves on partial duration series (PDS). The PDS are derived from event maxima series (EMS), and then fitted various distribution and finally determined the goodness-of-fit by the Anderson–Darling test. The best-fitted distribution is designated for predicting intensities related to the given duration and with a number of recurrence intervals. This procedure was repeated for eleven rainfall durations, from 5 to 240 min, at four stations of the Israel Meteorological Service.

Deidda (2010) introduced two objectives by using exceedances over a wide range of thresholds and proved by evaluating and comparing the performances of Monte Carlo samples drawn by GPDs with different shape and scale parameters and different discretizations.

Shinyie et al., (2012) estimated the Generalized Pareto Distribution (GPD) parameters using five methods namely the method of Moments (MOM), the probability weighted moments (PWM), the L-moments (LMOM), the Trimmed L-moments (TLMOM) and the Maximum Likelihood (ML) and the performance of the T-year return level of each estimation method was analyzed based on the RMSE measure obtained from Monte Carlo simulation of extreme rainfall events using the Partial Duration Series (PDS) method based on the hourly rainfall data of five stations in Peninsular Malaysia. In addition, they suggested the weighted average model, a model which assigns the inverse variance of several methods as weights, to estimate the value for T-year return period.

Pham et al., (2014) investigated the performance of the PDS/GP by setting different numbers of average peaks per year, with λ equal to 1, 2, 3, 4, and 5 across the North Island region in New Zealand. They found that the GP distribution is best defines with λ equal to 4 and 5 at almost all sites.

2.3 Application of Partial Duration Series for Regional Flood Frequency Analysis

Kinnison and Colby (1945) related flood frequency to the basin drainage characteristics to estimate the flood of any return period by fitting different distribution at an ungauged site. The USGS followed the study of Kinnison and Colby (1945) and started creating flood-frequency reports state by state based on the index-flood method.

At present, the index flood method is the most widely used regional flood frequency procedure (FSR, 1975; Hosking and Wallis, 1997; FEH, 1999; Castellarin et al., 2001; Brath et al., 2001; Sveinsson et al., 2001; Grover et al., 2002; Sveinsson et al., 2003; Lopez, 2004; Gaal et al., 2008). A homogeneity test is used to find a homogeneous region which leads to reducing the quantile estimation error. Dalrymple (1960) proposed such a test on homogeneity when the index flood approach of flood frequency analysis was introduced. The test, based on the assumption of an EV1 distribution, compares the variability of 10-year flood estimates, Q_{10}, from each site in the region with that expected supposing the differences between stations to be due to sampling error.

Wallis (1980) introduced the use of probability weighted moments (PWMs) in the index flood method. The technique calculates the PWMs at each site in a region from the standardized annual flood data and then the weighted regional average dimensionless PWMs are used to compute the dimensionless average growth curve. To obtain the T-year flood at a specific site, the dimensionless T-year growth curve X_T is multiplied by the at-site mean flow information, i.e., by the index flood. This technique was subsequently adopted as a viable way to estimate design floods and was further studied by, among others, Greis and Wood (1981, 1983); Lettenmaier et al. (1987); Stedinger and Lu (1995) and Hosking and Wallis (1997).

Mosley (1981) observed that cluster analysis of data describing the flood hydrology of selected New Zealand catchments was an attempt to identify regions in which catchments have a similar hydrologic regime. For the South Island, four regions in which the catchments were hydrologically more similar to each other than to catchments elsewhere were identified, but no similarly discrete regions could be identified in the North Island. It was apparent that where a number of factors were equally important in controlling hydrologic regime, a complex mosaic

of hydrologically homogeneous areas results, and no broad-scale regions could realistically be identified. Cluster analysis did not entirely eliminate subjective decisions, but greatly facilitated interpretation of a data set.

Tasker (1982) stated that data splitting was used to compare methods of determining "homogeneous" hydrologic regions by taking data from 221 stations in Arizona, USA. The methods used the complete linkage algorithm for cluster analysis and computed weighted average estimates of hydrologic characteristics at ungauged sites.

National Institute of Hydrology (1985) carried out a regional unit hydrograph study for Narmada basin based on Clark's approach. Here, the Clark model was derived for each of the sub-basins of Narmada basin using HEC-1 package. A regional value of $R/(t_c+R)$ along with the graphical relationship was used to estimate the parameters of the Clark model for the ungauged catchments of the Narmada basin.

Lettenmaier et al., (1987) explored the performance of index flood estimators with regions that exhibit various degrees of heterogeneity. Different variants of the Generalised extreme value GEV distribution with the PWM estimation procedure were used to obtain the index flood quantile. The GEV/PWM index flood quantile estimator performed well and gave the smallest mean squared errors in comparison with other at-site or regional quantile estimators for mildly heterogeneous regions.

Cunnane (1988) reviewed twelve different methods of regional flood frequency analysis including well-known methods such as the USWRC (U.S Water Resources Council) method, different variants of index flood methods, Station year methods, Bayesian methods and the two-component extreme value (TCEV) method and the index flood using a regional algorithm based on PWMs was rated as the best one. It was also recommended that either the Wakeby or GEV distribution be used when floods were estimated by the index flood method.

In a report for World Meteorological Organization (WMO), Cunnane (1989) presented a detailed review of various issues related to flood frequency analysis up to that time including statistical properties of observed flood series, the modeling problem, methods of quantile estimation and methods of choosing between distributions. This report also summarized a worldwide survey of flood frequency methods up to mid1980's. It recommended that flood estimates be based on the joint use of at-site and regional data using an Index Flood method of quantile estimation with model parameters estimated by probability weighted moments (PWMs). The report also revealed that conventional goodness of fit tests was of little value in

the context of choosing between statistical distributions and that the EV1 and LN distributions were the most commonly used distributions worldwide. However, the report discouraged the use of Log Pearson Type 3 (LP3) distribution in general for flood frequency procedures.

Another major development in relation to regional flood frequency analysis can be regarded as the Region of Influence (ROI) approach, developed by Burn (1990). The ROI technique involves the identification of a region of influence. The regional approach based on L-moments and the formation of a group of stations using the ROI approach are now well established methods and have been applied in many recent flood studies (Zrinji and Burn, 1994; Zrinji and Burn, 1996; FEH, 1999; Castellarin et al., 2001; Cunderlik and Burn, 2002; Shu and Burn, 2004; Merz and Bloschl, 2004; Cunderlik and Burn, 2006b; Gaal et al., 2008).

Nathan and McMahon (1990) presented a detailed regionalization methodology that addresses the problems associated with the selection of an appropriate clustering technique, selection of catchment variables, the definition of homogeneous regions, and the prediction of group membership for other catchments whose group membership was otherwise unknown. The most suitable technique that they identified used multiple regressions to select and weight the most appropriate variables and then used cluster analysis to derive preliminary groupings, finally applying a multi-dimensional plotting technique to investigate further and refine the preliminary groupings.

Singh and Kumar (1991) carried out a study using the peak flood series data of hydro-meteorologically homogeneous region of Godavari basin Subzone 3f involving application of EVI (PWM) and GEV (PWM) methods based on i) on-site data, ii) on-site and regional data combined and iii) regional data alone. Homogeneity of the region was testing using USGS and Coefficient of variation based homogeneity test. From the study, it was concluded that GEV (PWM) approach using on-site and regional data in a combined form would provide estimates of flood peaks for different return periods with computationally less bias, and comparable root mean square error.

Lu and Stedinger (1992) formulated a significance test of homogeneity, based on the variability of normalized on-site GEV flood quantiles (X_{10}) estimated by L-moments and demonstrated that this test was more powerful than the Wiltshire's R-statistic test.

Vogel and Kroll (1992) presented a simple conceptual stream-aquifer model which was extended to a watershed scale and evaluated for its ability to approximate the low-flow behavior of 23 unregulated catchments in Massachusetts, USA. The conceptual watershed model was

then adapted to estimate low-flow statistics using multivariate regional regression procedures. Their results indicated that in central western Massachusetts, low-flow statistics were highly correlated with the product of the watershed area, average basin slope and base flow recession constant.

Fovell and Fovell (1993) used monthly rainfall and temperature data for recognizing climatic regions of the United States by using Hierarchical cluster (HC) in combination with principal component analysis (PCA).

Hosking and Wallis (1993, 1997) proposed two more homogeneity tests based on L-moment ratios such as using L-CV alone (H1) and using L-CV & L-skewness jointly (H2). Both tests measure the sample variability of the L-moment ratios among the samples in the pooling group and compare it to the variation that would be expected in a homogeneous pooling group. The variation was estimated through repeated simulations of homogeneous regions with samples drawn from a four parameter kappa distribution whose parameters were estimated from L-CV, L-skewness and L-kurtosis of the region's data. They recommended using the H1 over the H2 statistic as they found that the heterogeneity based on L-CV had better power to discriminate between homogeneous and heterogeneous regions.

Madsen et al., (1994) employed the partial duration series method as an alternative to the traditional non-parametric approach in the modeling of extreme rainfalls. In order to obtain an estimation procedure at non-monitored sites and to improve at-site estimates, a regional Bayesian approach was adopted. The empirical regional distributions of the parameters in the Partial Duration Series model were used as prior information. The application of the Bayesian approach was derived in the case of both exponential and generalized Pareto distribution exceedances. Finally, the aspect of including economic perspectives in the estimation of the design events was briefly discussed.

Birikundavyi et al., (1997) analysed flood estimation techniques at both at-site and a regional context using partial duration series in the province of Ontario, Canada. They also analysed the problem of selection of the threshold. They considered Poisson distribution to define the occurrence of floods and the generalized Pareto distribution for flood magnitudes.

Madsen et al., (1997) compared two different models such as partial duration series (PDS) and Annual maximum series (AMS), for analyzing extreme hydrologic events. Then the performance in terms of return period events estimator is evaluated using different parameter estimation methods. They concluded that in the case of PDS model with negative shape

parameter and for AMS model with moderately positive shape parameters, the Method of Moments (MOM) estimation gives good results whereas, Maximum Likelihood (ML) estimation gives good results in the case of the PDS model for large positive shape parameters.

A publication by FEH (1999) studied a range of pooling group sizes and its obvious impact on adoption of the 5T rule, namely that the total number of station years of data to be included when estimating the T-year flood should be at least 5T. An advantage of the region of influence method was that in the estimation of a regional growth curve, each site can be weighted according to its closeness to the site of interest.

Kumar et al., (1999) developed regional flood frequency curves by fitting the L-moment based generalized extreme value distribution to annual maximum peak flood data of small-to-medium size catchments of the seven hydrometeorological subzones of Zone 3 of India. They developed the regional flood frequency curves for each subzone together with on-site mean annual peak floods for both gauged and ungauged catchments.

Adamowski (2000) found that the analysis of annual maximum (AM) flood series had revealed unimodal and multimodal probability density functions for floods in the Provinces of Ontario and Quebec, Canada and divided the region into nine homogeneous regions having similar flood generating nature. Then a regional relationship was developed using nonparametric analysis on the AM and PDS.

Brath et al., (2001) analyzed three indirect techniques for index flood estimation in order to evaluate their applicability and effectiveness. This analysis was based on both statistical and conceptual approaches, in 33 hydrometric stations of northern-central Italy. The results showed that the statistical model was best when compared to the physically-based models. Finally, the results highlighted that direct estimation techniques could be advisable for catchments with intrinsic geomorphoclimatic properties.

Castellarin et al., (2001) evaluated the relative performance of four hydrological similarity measures for regional frequency analysis in Northern-Central Italy and computed their performance based on Monte Carlo experiment. From the results, it was found that the similarity measures based on seasonality indexes were effective for estimating extreme flow quantiles for the study area.

Lim et al., (2003) examined the flood records of more than 23 gauged river basins in Sarawak, Malaysia, using an index-flood estimation procedure based on L-moments. Two homogeneous

regions were identified. The Generalized Extreme Value and the Generalized Logistic distributions were found to describe the distribution of extreme flood events appropriately within the respective regions. A regional growth curve was subsequently developed for each of the regions for the estimation of design floods in ungauged basins.

Jingyi and Hall (2004) investigated the homogeneous region in the Gan-Ming river basin of China by using different clustering techniques include K-mean (KM), fuzzy C-mean (FCM), hierarchical clustering (HC), and Kohonen self-organizing features map. They also applied the appropriateness of the Kohonen map for finding the number of clusters and sites selected.

Parida, (2004) attempted to identify homogeneous regions using physical, hydrological and meteorological attributes that are responsible for flood generation in flood stricken area of eastern India. Euclidean distance was used to classify basins into regions which yield minimum partitioning error. The identified regions were subjected to several homogeneity tests which revealed that the method throws a promise for use in regional flood frequency analysis to obtain reliable results.

Chowdhury, (2005) attempted to evolve a suitable methodology for determining the floods of given return periods in ungauged catchments of river Mahanadi in Eastern India based on Regional Flood Frequency Analysis procedures. The Index Flood Method and Multiple Regression Techniques have been used. The flood data of 16 gauging sites have been utilized to evolve the regional relationships. The flood value results obtained by the developed relationships of two methods when compared with the flood values derived from best-fitted frequency distribution indicated that relationship developed by Multiple Regression Technique was quite suitable for Mahanadi basin.

Kumar and Chatterjee (2005) examined the regional flood frequency analysis using data of 13 stream flow gauging sites of the North Brahmaputra region of India. They found that General extreme value (GEV) distribution was a robust distribution for the study area and developed a regional relationship between mean annual peak flood and catchment area for quantile estimation for ungauged sites.

Trefry et al., (2005) worked on regional frequency analysis using PDS series for the state of Michigan, USA. They found that the PDS/GP model performed well with λ ranging from 2.2 to 4.07 and hence, $\lambda_R = 2$ was selected for regional modeling.

Goel and Arya (2006) developed a dynamic flood frequency model for estimation of parameters of stochastic rainfall models including regionalisation of rainfall parameters. The framework of methodology for dynamic flood frequency models or derived flood frequency distribution (DFFD) models consisted of the following three major components: (1) Stochastic rainfall model (2) Infiltration model and (3) Effective rainfall–runoff model.

Rao et al., (2006) tested, the effectiveness of Fuzzy cluster analysis (FCA) for regionalization by using annual maximum flow data from the watersheds in Indiana, USA. The effectiveness of several fuzzy cluster validation measures in determining optimal partition provided by the FCA was also addressed.

Norbiato et al., (2007) focused on flash floods in the eastern Italian Alps due to heavy rainfall and large spatial variation. They applied regional frequency analysis using the index variable method and L-moments to examine short duration maximum precipitation for the Friuli-Venezia Giulia region, in north-eastern Italy, which included the storm location and found that the Kappa distribution may be useful. Then various severity graphs were established to visualize the return periods and their variability for different rainfall durations within the storm.

Stambuk et al., (2007) investigated possible application of the Kohonen self-organizing maps (SOM) to social sciences data clustering and compared the results of the procedure to the Principal component analysis (PCA) and Hierarchical Cluster (HC) methods.

Viglione et al., (2007) compared four homogeneity tests through the determination of the power associated with the tests using Monte Carlo simulation experiment. The first two of these tests are those of Hosking and Wallis (1997), who proposed H1 and H2, and the other two, introduced by the authors, were based on the k sample Anderson-Darling test and Durbin & Knott test. They concluded that the H2 as a homogeneity test lacked power. They further concluded that the H1 test should be preferred when skewness is low while the Anderson-Darling test should be used for more skewed regions, preferably those with L-skewness greater than 0.23.

Eslamian and Biabanaki (2008) worked on regionalization by applying cluster analysis and Andrews plot (AP) in the Kharkeh basin, Iran of low flow. Cluster analysis was used to classify the data in order to capture a diversity of factors, and the K-means algorithm was specified for assigning stations to a cluster.

Beaulieu et al., (2009) presented an intercomparison of eight statistical tests to detect inhomogeneities in climatic data of the province of Quebec, Canada. They found that none of these methods was efficient for all types of inhomogeneities, but some of them performed substantially better than others: the bivariate test, the Jaruskova's method, and the standard normal homogeneity test.

Chavoshi and Soleiman (2009) worked on regional flood frequency analysis by using conventional cluster analysis and fuzzy logic theory for 70 catchments in northern Iran which was applied both in low flow and rainfall analysis.

Saf (2009) investigated hydrologically homogeneous regions through regional flood frequency estimates for 47 gauged sites in the West Mediterranean River Basins in Turkey, using an index flood method with L-moments. In the study three subregions were found and based on L-moments goodness-of-fit statistic, the Pearson type III distribution as identified as the best-fit distribution for the Antalya and Lower-West Mediterranean subregions, while the Generalized Logistic distribution for the Upper-West Mediterranean subregion. Then Monte Carlo simulation was used to evaluate the accuracy of the quantile estimates on the basis of the relative root-mean-square error and relative bias.

Borujeni et al., (2010) analyzed the peak floods, observed in North-Karoon basin, Iran using five distributions and estimated the parameters using the L-moment method. They found 5 homogeneous regions out of 7 study sites, and lognormal distribution was identified as the most appropriate distribution in the homogeneous study region. The regional peak flood estimates for each return period were obtained based on this distribution.

Das et al., (2010) examined how successful a common method of identifying pooling group membership was in selecting groups that actually were homogeneous based on annual maximum series obtained from 85 Irish gauging stations. Each station had its own unique pooling group selected by use of a Euclidean distance measure. The results were also compared with the heterogeneity measures H1 and H2 and found that 27 pooling groups were heterogeneous. These groups were further examined with the help of box-plots. From the results, it was concluded that it was not sure to identify perfectly homogeneous groups.

Kar et al., (2010) investigated the partition of the Mahanadi basin in Eastern India into homogeneous regions by applying different clustering techniques by using fewer but influential variables. Principal component analysis was used for finding important variables. The results obtained from different clustering techniques were useful for selection of a number of sites

present in a particular cluster. Homogeneity test was carried out by using the regional L-moment algorithm and to find a suitable frequency distribution. Finally, an index flood method was applied, and the obtained results were compared with the earlier studies of flood frequency in this basin.

Saf (2010) analysed how outliers affect the identification of regional probability distributions using L-moments method in a region of the Menderes River Basins in Turkey. Through various tests two sub-regions viz. the Upper-Menderes and Lower-Menderes sub-regions were analyzed. Based on the L-moments goodness of statistic criteria, the generalized extreme value distribution was determined as the best-fit distribution for both the regions. Here the generalized extreme value distribution was found to be the best-fit distribution for the Upper-Menderes sub-region and the Pearson Type 3 distribution was found to be the best for the Lower-Menderes sub-region based on a robust measure. It was concluded that the homogeneous region determined from the robust discordancy measure was more accurate than the region identified using the classical robust measure.

Yang et al., (2010) presented a method for regional frequency analysis and spatio-temporal pattern characterization of rainfall-extreme regimes in the Pearl River Basin (PRB) in China using the L-moments approach along with stationarity test and serial correlation check. From the results, it was found that in the Basin which was divided into six regions, the Generalized Normal (GNO), Generalized Logistic (GLO), Generalized Extreme Value (GEV), and Pearson Type 3 (PE3) distributions fit well for different regions. Then the quantiles were estimated by Monte Carlo simulation which gave reliable results for the return periods of less than 100 years. Also, they observed the high precipitation at Guilin region of Guangxi Province and Fogang region of Guangdong Province which was responsible for flood disasters in the regions.

Malekinezhad et al., (2011) compared two regional flood frequency methods including index-flood and multiple-regression analyses based on L-moments in the Namak-Lake basin in central Iran. To do so they delineated homogeneous regions using cluster analysis, checked the homogeneity and fitted the distribution. From the results, it was concluded that for the basin divided into three regions,the generalised extreme value distribution was the best-fit distribution. After that to evaluate the performance of both the methods, relative root mean square error (RRMSE) measure was applied.

Gebregiorgis et al., (2013) discussed on regionalization of the Blue Nile River Basin (BNRB) in Africa by using statistical techniques and described the selection of best-fit distribution models to estimate the flood frequency in the basin. They found five homogeneous regions and

fitted 14 different distribution with different parameter estimation methods. A unique regional flood frequency curve was developed for each region to estimate the flood quantiles of the ungauged part of the basin.

Pham et al., (2014) developed a unique regional value of λ_R of PDS across the North Island region of New Zealand based on both on-site and regional PDS series. This value estimated the most consistent quantiles for both small and large return periods.

2.4 Flood Forecasting using Artificial Neural Network and Neuro Fuzzy Inference System

Flood prediction and its mitigation or management is one of the greatest challenges facing the world today. Floods have become more frequent and severe due to effects of global climate change and human alterations of the natural environment. All flood forecasting systems serve specific purposes and in most cases they are designed to prevent, minimize, or mitigate people's suffering and to limit economic losses. The forecast of flooding would benefit greatly from the use of hydrological models, which are designed to simulate flow processes of surface or subsurface water. Flood models are mainly used in flood forecasting and early warning systems. Both systems require the reliable real-time hydro-meteorological data and lag time. The utility of forecasting depends largely on the relationship between the desired lead time at that point and the lag time of the hydrological response (Lettenmaier and Wood, 1993; Werner et al., 2005).

Karunanithi et al., (1994) demonstrated how a neural network can be used as an adaptive model synthesizer as well as a predictor using a constructive algorithm called the cascade-correlation algorithm, which was applied to the flow prediction of the Huron River at the Dexter sampling station, near Ann Arbor, Michigan, USA. They found the performance of the network based on the algorithm were proficient in adjusting their complexity to match changes in the flow history.

Minns et al., (1996) analysed a series of numerical experiments, which generate the flow data from synthetic storm series directed through a conceptual hydrological model consisting of a single nonlinear reservoir. They demonstrated the closeness of fit that can be achieved with such data sets using Artificial Neural Networks (ANNs).

Thirumalaiah and Deo (1998) highlighted the use of the neural networks in real-time forecasting of water levels at a given site continuously throughout the year based on the same

levels at some upstream gauging station and/or using the stage time history recorded at the same site. The network was trained by using three algorithms, namely, error back propagation, cascade correlation, and conjugate gradient. Then they compared the training results with each other and verified with untrained data.

Campolo et al., (1999) developed a neural network model to analyze and forecast the behavior of the river Tagliamento, in Italy, during heavy rain periods. The model used the distributed rainfall information to predict the water level of the river. From the result, it was observed that the model with a 1- hour time horizon would predict accurately whereas with an increase of the time horizon the prediction accuracy is decreased. Finally it was concluded that the performance of the model would remain satisfactory up to 5 hours.

Zealand et al., (1999) investigated the value of Artificial Neural Networks (ANNs) for short-term forecasting of streamflow in the Winnipeg River system in Northwest Ontario, Canada. They found that a very close fit was obtained during the training phase, and the ANNs developed consistently outperformed a conventional model during the testing phase for all of the four forecast lead-times.

Coulibaly et al., (2000) introduced an early stopped training approach (STA) to train multilayer feedforward neural networks (FNN) by considering the hydrological time series from the Chute-du-Diable hydro system in northern Quebec (Canada) for real time flood forecasting. The performance of the model was compared with a statistical model and an operational conceptual model to test the real-time forecast accuracy and it was found that the proposed method was effective for improving prediction accuracy.

Spokkerreff (2000) mentioned about the application of statistical model till 1995 for river Rhine in Germany. The hydrological years 1994 and 1995 were characterized by two extraordinary extreme floods in the basins of the River Meuse and the River Rhine. Water levels were measured with return periods of more than 100 years, considerable damage occurred and as a precaution over 2,00,000 people had to be evacuated. Both events showed the importance of reliable forecasts with a sufficient forecast period. Until then forecasts for the Rhine River were carried out with a statistical model, allowing a reliable two-day forecast for the Lobith gauging station on the German/Dutch border. To extend this forecast period, a new flood forecasting model was developed. The first operational use of the new model during some minor floods at the beginning of 1999 showed reliable results for the three-day forecast and considerable improvement for the four-day forecast.

Valenca et al., (2000) presented a Fuzzy Neural Network model for inflow forecast for the Sobradinho Hydroelectric power plant, part of the Chesf (Companhia HidrelCtrica do Siio Francisco-Brazil) system. The model was implemented to forecast monthly average inflow on a one-step-ahead basis. The Fuzzy Neural Network model was found to provide a better representation of the monthly average water inflow forecasting, than the models based on Box-Jenkins method.

Chang et al., (2001) stated that a counter propagation fuzzy-neural network (CFNN) is the fusion of a neural network and fuzzy arithmetic. They used the streamflow and precipitation data of the upstream of the Da-cha River, in central Taiwan, to evaluate the CFNN rainfall-runoff model and compared their results with the ARMAX. They found that the CFNN rainfall-runoff model was superior and reliable.

Dawson et al., (2001) considered the application of artificial neural networks (ANNs) to rainfall-runoff modeling and flood forecasting. They proposed a template in order to assist the construction of future ANN rainfall-runoff models. Finally, it was suggested that research might focus on the extraction of hydrological 'rules' from ANN weights, and on the development of standard performance measures that penalize unnecessary model complexity.

Hundecha et al., (2001) developed fuzzy rule-based routines to simulate the different processes involved in the generation of runoff from precipitation and validation of the model was done on a rainfall-runoff analysis for the Neckar River catchment, in southwest Germany.

Xiong et al., (2001) introduced, the first-order Takagi–Sugeno fuzzy system and explained as the fourth combination method [besides other three combination methods tested earlier, i.e. the simple average method (SAM), the weighted average method (WAM), and the neural network method (NNM)] to combine together the simulation results of five different conceptual rainfall-runoff models in a flood forecasting study on eleven catchments. Due to the simplicity and efficiency, of the first-order Takagi–Sugeno method it was recommended for use as the combination system for flood forecasting.

Dolling et al., (2002) presented monthly streamflow prediction using artificial neural networks (ANN) on mountain watersheds. From the results, it was found that the spring and summer monthly streamflows could be adequately represented, improving the results of calculations obtained using other methods which had significant benefits for the optimal use of water resources for irrigation and hydroelectric energy generation.

Shamseldin et al., (2002) applied a Multi-Layer Feed-Forward Neural Network (MLFFNN) in the context of river flow forecast combination, where a number of rainfall-runoff models were used simultaneously to produce an overall combined river flow forecast. They used five neuron transfer functions, namely, the logistic function, the bipolar function, the hyperbolic tangent function, the arctan function and the scaled arctan function and found that the logistic function would yield the best model forecast combination performance.

Sudheer et al., (2002) presented a new approach to designing the network structure in an artificial neural network (ANN) based rainfall-runoff model. Their method utilized the statistical properties such as cross-correlation, auto-correlation and partial-auto-correlation of the data series in identifying a unique input vector that would best represent the process for the basin, and a standard algorithm for training. The methodology has been validated using the data for a river basin in India. The results of the study were highly promising and indicated that they could significantly reduce the effort and computational time required in developing an ANN model.

Campolo et al., (2003) presented a real- time flood forecasting model for the Arno basin in the Tuscany region of Italy under low flow conditions to predict the water-level evolution based on the artificial neural network. They found that the prediction of water level would remain accurate within a forecast time ahead of 6 h, and it would increase for each time ahead of prediction, as the flow rate increases, signifying that the model was mostly appropriate for flood forecasting purposes.

Jain et al., (2004) developed a new approach using real-coded genetic algorithms (GAs) do not use any coding of the problem variables, instead they work directly with the variables and a new class of models to train ANN rainfall-runoff models using the daily rainfall and streamflow data from the Kentucky River watershed. They found that the results obtained from the models using real-coded GA were able to predict daily flow more accurately. Also, they found that the grey box models performed better than the purely black box type ANN rainfall-runoff models.

Nayak et al., (2004) presented the application of an adaptive neuro-fuzzy inference system (ANFIS) to hydrologic time series modeling to model the river flow of Baitarani River in Orissa state, India. The results showed that the ANFIS forecasted flow series would preserve the statistical properties of the original flow series. The results were highly promising, and comparative analysis suggested that the proposed modeling approach outperforms ANNs and other traditional time series models in terms of computational speed, forecast errors, efficiency, peak flow estimation, etc.

Solomatine and Xue (2004) stated that advances in data-driven modeling have improved the accuracy of forecasts made using physically based models. They have drawn attention to the innovative use of such techniques for flood forecasting in rivers.

Chau et al., (2005) employed two hybrid models based on recent artificial intelligence technology, namely, the genetic algorithm based artificial neural network (ANN-GA) and the adaptive-network-based fuzzy inference system (ANFIS), for flood forecasting in a channel reach of the Yangtze River in China. They observed that the performance of both hybrid algorithms was more accurate than the linear regression model.

Nayak et al., (2005a) analysed the potential of fuzzy computing based rainfall–runoff model in real time flood forecasting by developing a model for forecasting the river flow of Narmada River in India. They demonstrated that fuzzy models can take advantage of their capability to simulate the unknown relationships between a set of relevant hydrological data.

Nayak et al., (2005b) explored the potential of the neuro-fuzzy computing paradigm to model the rainfall-runoff process for forecasting the river flow of Kolar basin in India. They found that the forecasts by the neuro-fuzzy model at higher lead times (up to 6 hours) were found to be better than those from the neural network model or the fuzzy model.

Sudheer (2005) discussed a perturbation analysis for determining the order of influence of the elements in the input vector on the output vector through a case study of a river flow model developed for the Narmada River, India. They found that each variable in the input vector influenced the shape of the hydrograph in different ways. However, the magnitude of the influence could not be clearly enumerated by this approach.

Chang et al., (2006) used the adaptive network-based fuzzy inference system (ANFIS) to build a prediction model for reservoir management through a case study of the Shihmen reservoir, Taiwan by considering typhoon (i.e., cyclone) and heavy rainfall events with 8640 hourly data sets collected in the past 31 years. They developed two ANFIS models: one with human decision as input, another without human decision input and concluded that the model with human decision as input variable would provide high accuracy and reliability for reservoir water level forecasting.

Lohani et al., (2006) investigated the potential of Takagi–Sugeno (TS) fuzzy inference system for modeling stage–discharge relationships of various gauging stations in Narmada river system, India. The results showed that the TS fuzzy modeling approach was superior to the

conventional and artificial neural network (ANN) based approaches. Also, the approach was able to model the hysteresis effect (i.e.,loop rating curve) more accurately than the ANN approach.

Aqil et al., (2007) examined the advantages of artificial neural networks and neuro-fuzzy system in the continuous modeling of the daily and hourly behavior of runoff. From the results, it was found that the neuro-fuzzy model performed better than both the Levenberg–Marquardt-FFNN and the Bayesian regularization-FFNN.

Chang et al., (2007) presented a systematic investigation of three common types of artificial neural networks (ANNs) for multi-step-ahead (MSA) flood forecasting such as multi-input multi-output (MIMO), multi-input single-output (MISO) and serial-propagated structure for two watersheds in Taiwan. From the results, it was observed that the MIMO was less accurate, whereas both MISO and serial-propagated neural networks were capable of performing accurate short-term forecasting. For long-term forecasts, only the serial-propagated neural network could provide satisfactory results in both watersheds.

Mukerjee et al., (2009) carried out flood forecasting at Jamtara gauging site of the Ajay River Basin in Jharkhand, India using an artificial neural network ANN model, an adaptive neuro-fuzzy interference system ANFIS model and an adaptive neuro-GA integrated system ANGIS model. They found that ANGIS model with the same input dataset predicted flood events with maximum accuracy and between ANFIS and ANN models, ANFIS predicted better in most of the cases.

Kar et al., (2010) attempted to develop a workable forecasting system for the downstream catchment of Mahanadi River in Eastern India by taking the concurrent flood peaks for 12 years based on both statistical method and ANN based approach. A comparison between both methods were tested, and it was found that the ANN methods were better beyond the calibration range over statistical method, and the efficiency of either method would reduce as the prediction reach was extended.

Kar et al., (2012) developed a flood forecasting model for Ayeyarwady River Basin of Myanmar applying ANN multilayered feed-forward network along with the Takagi-Sugeno (TS) fuzzy inference model, to forecast the stage for 1 to 4 days in advance. They found that T-S (Takagi-Sugeno) fuzzy model (Takagi and Sugeno 1985) would perform better than the MLFF (i.e., multilayered feed-forward) network.

2.5 Flood Inundation Mapping and 1-D Hydrodynamic Modeling

The airborne and satellite synthetic aperture radar (SAR) is considered for measuring flood extent due to its cloud-penetrating day and night capacity. Ormsby and Blanchard (1985) as well as Pultz and Crevier (1997) supported some of the earliest experimental work on SAR response to flooded vegetation using X-band, C-band and L-band imagery and concluded that response would depend on wavelength, plant volume and the geometry of the inundated vegetation.

In India, the area affected from floods has increased from 2.29 to 4.94 million hectares from the year 1953 to 2000. During this period, the loss of human lives increased from 37 to 2345 and the monetary damages increased from about 11 to 295 million US dollar as per studies by Central Water Commission (CWC, 1997).

Bates et al., (1997) mentioned the further development of two-dimensional finite element models of river flood flow. They applied the two-dimensional finite element model to the Missouri river in Nebraska, USA with the integration of hydraulic modeling and remote sensing.

Shafiee, et al., (2000) investigated the capability of temporal and multimode Radarsat data for monitoring of flood. They have overlaid the Radarsat images with Landsat-5 TM to produce the flood mapping and concluded that SAR images have potential and capability to monitor and map the flood event.

Han et al., (1998) and Chang et al., (2000) have also reported 1-D, 2-D coupled modeling of river floodplain flow. The models have used a full dynamic equation for the channel flow and for the two-dimensional floodplain flow; a diffusion wave approximation was utilized.

Hydrologic engineering Center- River Analysis system (HEC-RAS) and GIS technologies are integrated to obtain scientifically derived information that has been quantified as effective in simulating, identifying and analyzing flood events in a geo-spatial environment by Shamsi (2002). This helps in visualizing flood simulations, and can view the spatial impact of various scenarios along with the critical locations to assess the vulnerability of the area towards a flood event efficiently.

Anderson (2000), Robayo et al., (2004) and Knebl et al. (2005) concluded that flood inundation modeling involves hydrologic modeling to estimate peak flows from storm events, hydraulic

modeling to estimate water surface elevations, and terrain analysis to estimate the inundation area.

Mostly, studies have applied hydraulic and hydrological models for simulating flood runoff and runoff in low-lying flood-prone areas, in order to provide flood risk assessment information on the probability of flood occurrence, magnitude of the event, location and depth of the inundation for flood management as observed by Booij (2005).

Goel et al., (2005) presented a technique for preparation of flood hazard maps which include the development of DEM (digital elevation model) and simulation of flood flows for different return periods.

Mathematical models for flood simulation solve a set of governing equations and provide specific information on flood characteristics as observed by Haile (2005).

Wright et al., (2008) presented a methodology for using remotely sensed data to both generate and evaluate a hydraulic model of floodplain inundation for a rural case study in the United Kingdom viz, Upton-upon-Severn.

Zheng et al., (2008) developed a distributed model for simulating flood inundation integrating with rainfall-runoff processes using Shuttle Radar Topography Mission (SRTM)-DEM data and some remote sensing datasets in the environment of GIS for Maruyama River basin, Japan. Simulated results in the Maruyama River basin demonstrated an acceptable agreement with the flooded area observed.

Patro et al., (2009) used a coupled 1-D and 2-D hydrodynamic model, viz, MIKE FLOOD to simulate the flood inundation extent and flooding depth in the delta region of Mahanadi River basin in India. They used SRTM-DEM to prepare a bathymetry of the study area and provided as an input to the 2D model, MIKE 21. Using lateral links in MIKE 11 and MIKE 21 models flood inundated area was obtained. Results were compared with actual inundated area obtained from IRS-1D WiFS image.

In Aaron Cook (2009), the effect of topography, geometric configuration and modeling approach on two different study areas have been addressed. It was found that inundation area would decrease as the resolution and vertical accuracy of topographic data increases. FESWMS (Finite Element Surface Water Modeling System) was used by them which generated less inundation area as compared to HEC-RAS. Variation in inundation extent with respect to

resolution is less in FESWMS then HEC-RAS. All its findings conclude that 2D modeling approach is more realistic than 1D approach.

Bhatt et al., (2010) discussed the operational use of remote sensing technology for near real-time flood mapping, monitoring of Kosi river floods in Nepal, India and the satellite-based observations made for the Kosi river breach.

Samarasinghe et al., (2010) derived flood extent from the flood extent obtained for the 50-year rainfall using HEC-HMS and HEC-RAS in Kalu-Ganga River, Sri Lanka.

Bashir et al., (2010) generated flood hazard map for Nullah Lai in Rawalpindi, Pakistan using HEC-RAS and HEC-GeoRAS hydrological models with GIS. They found a relationship between inundation depth and specific discharge value.

Shaohong et al., (2010) developed a real-time flood monitoring system that would permit integrated handling of hydrological data coming from a wireless monitoring network. They obtained water surface elevation using hydrological data and spatial position information using spatial analysis technology in GIS software. Then, flood area information was analyzed by deduction of water surface elevation in the digital elevation model.

According to Orok (2011), that flood risk maps should be able to identify the areas that are most vulnerable to flooding and estimate the number of people that will be affected by floods in a particular area.

Adnan et al. (2012) carried out bathymetry mapping based on remotely sensed imagery coupled with ancillary datasets for River Kelantan, Malaysia using the hydraulic model HEC-RAS. Predicted flood inundation extent using HEC-RAS was compared to flood extent predicted from a RADARSAT image. The accuracy assessment was applied to identify spatial variation in the error among three areas (i.e. upstream, midstream and downstream).

Preparation of flood maps would provide valuable information for managers and experts to reduce flood damages as observed by Hassanpour et al. (2012).

A large scale flood inundation forecasting was also carried out by using Lisflood-fp model, by Neal et al., (2012). This study was carried out on Lower Zambezi River to demonstrate current flood inundation forecasting capabilities in large data-scarce regions. Here they used newly developed sub-grid channel scheme to describe river network. The model evaluation showed

that simulated flood edge cells were within a distance of one to two model resolutions when compared to an observed flood edge, and inundation area was accurate by about 86% on an average.

Hydro-meteorological catastrophes cannot be totally evaded, but the impacts and after effects can be managed by developing effective risk reduction strategies through the application of latest geospatial tools and decision support systems as observed by Sadiq et al., (2014).

As described above in this Chapter all the relevant literature on floods related to flood frequency analysis by partial duration series, regional flood frequency analysis by partial duration series, flood forecasting using ANN and neuro fuzzy inference system as well as flood inundation mapping/ 1-D hydrodynamic modeling has been reviewed.

CHAPTER 3
THE STUDY AREA AND DATA COLLECTION

This chapter gives a description of the study area, the Mahanadi River basin, and the data collected from different sources for the study and analysis of the present work. The River Mahanadi is the 8th largest basin of the country originates from the Amarkantak Hills of the Bastar Plateau near Pharasiya village in Raipur district of Chhattisgarh and the data used are daily rainfall, daily discharge data, water level, topographic information, soil characteristics and land use changes.

3.1 THE STUDY AREA

3.1.1 Overview of Basin

The Mahanadi River is one of the major inter-state east flowing rivers in peninsular India after the Godavari with respect to the water potential and flood producing capacity. It originates at an elevation of 442 m above mean sea level and lies within geographical coordinates of 80°30'E–86°50'E longitude and 19°20'N–23°35'N latitude (Figure 3.1) and covering major parts of Chhattisgarh, Odisha and comparatively smaller portions of Jharkhand, Maharashtra and Madhya Pradesh (Figure 3.2). The basin is largely divided into four parts such as Central table land, Eastern ghats, Northern plateau and Coastal plain. A large reservoir named Hirakud is situated at the center of the catchment with the drainage are of 83000 km^2 out of which about 65580 km^2 of area is lying in Odisha. The total length of the river from its origin to confluence in the Bay of Bengal is about 851 km, of which 357 km is in Chhattisgarh, and 494 km is in Orissa.

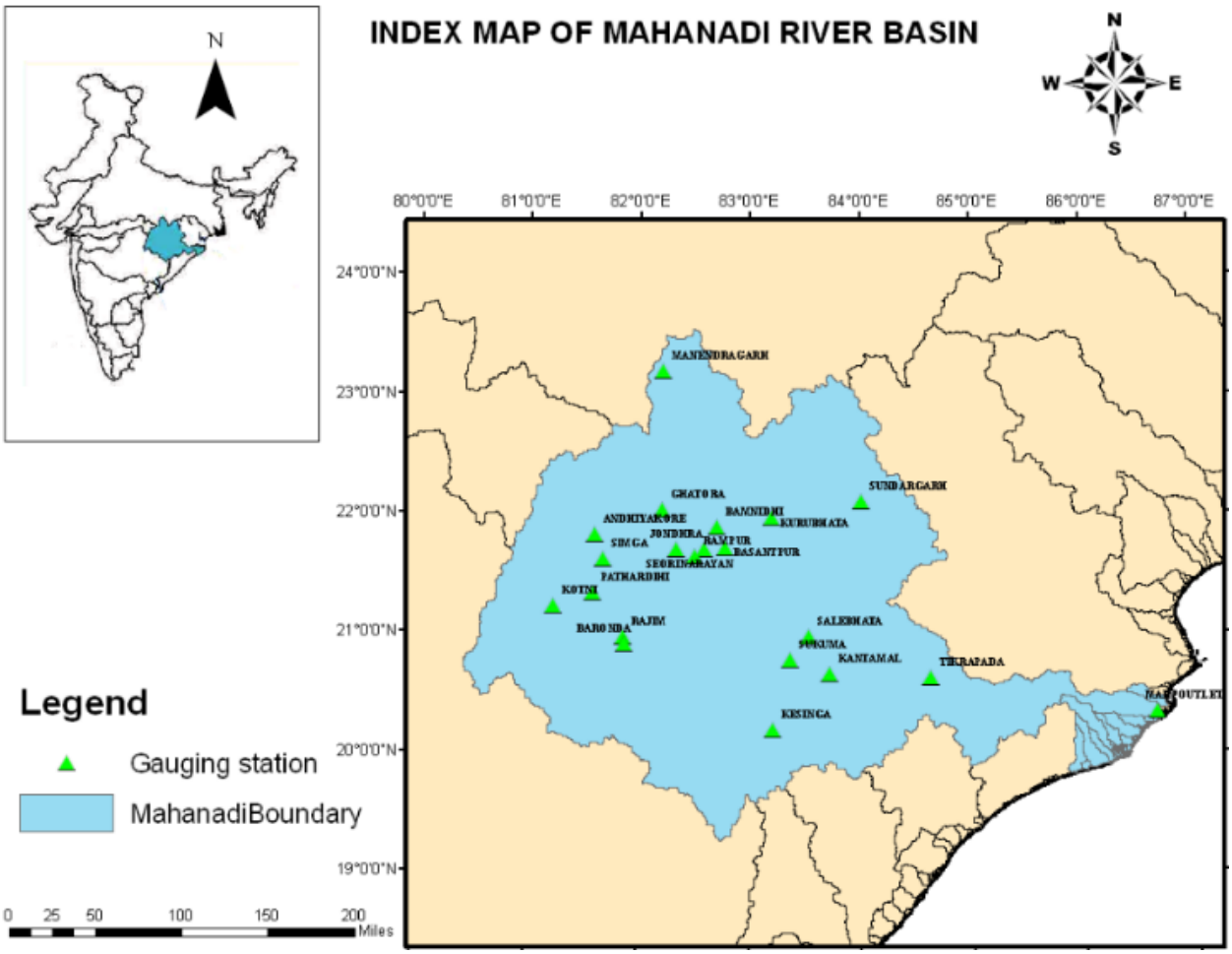


Figure 3.1: Study area and discharge sites location in of Mahanadi River

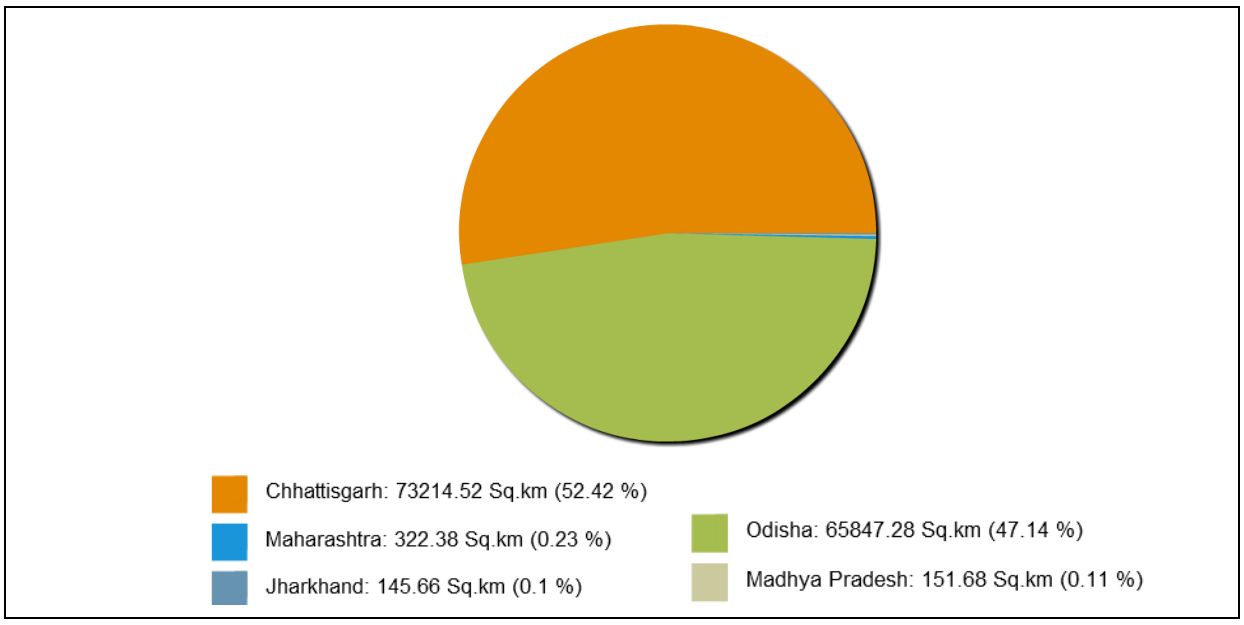


Figure 3.2: State-wise share of basin area of Mahanadi River Basin

Upper Mahanadi sub-basin is drained mainly by the Seonath, the Arpa, the Kurung and the Sakri rivers. Middle Mahanadi sub basin is comprised of the Mahanadi, the Jonk, the IB, the Bhedan and the Mand rivers and the Lower Mahanadi sub-basin covers southern and coastal part of the basin and it is drained by the Ong, the Tel, the Hati and the Daya rivers. It traverses a distance of 320km up to Naraj Barrage and after Naraj the river splits into several distributaries such as Kathjodi, Birupa, Kuakhai, Chitrotpala, Luna, Karandia, Paika and Sukapaika. Kathjodi is a large distributary and branches into Kathjodi, Surua, Biluakahi, Devi, Kandal, Taunla which again join together and fall into the Bay of Bengal after entering Puri district. The Delta formed by numerous distributaries of the Mahanadi and the Brahmani, is one of the largest deltas in India. During the course of Delta formation, some islands have been formed between various channels and those islands are subjected to continual flooding during the monsoon due to spill over beyond the channels.

3.1.2 Topography

The Mahanadi basin has varying topography with the highest elevation found in northern hills and the lowest elevation in coastal reaches as shown in Figure 3.3. The maximum elevation observed is 1254 m in the steep hilly terrain of Mahanadi basin. The Upper Mahanadi sub-basin with its predominantly hilly terrain in its northern upper part has elevation ranging from 750-1000 m. The central flank of the upper Mahanadi which is drained by Seonath River is a plain area having elevation range of 200 to 300 m surrounded by higher hills on its west having a height between 300 and 400 m. The middle Mahanadi sub-basin has high hilly terrain in its north-eastern stretch. This part has the highest elevation which falls between 750-1000 m. The Coastal plain stretching over the districts of Cuttack and Puri covers the large delta by Mahanadi and elevation decreases towards this deltaic stretch reaching up to 10-50 m (Source: www.india-wris.nrsc.gov.in).

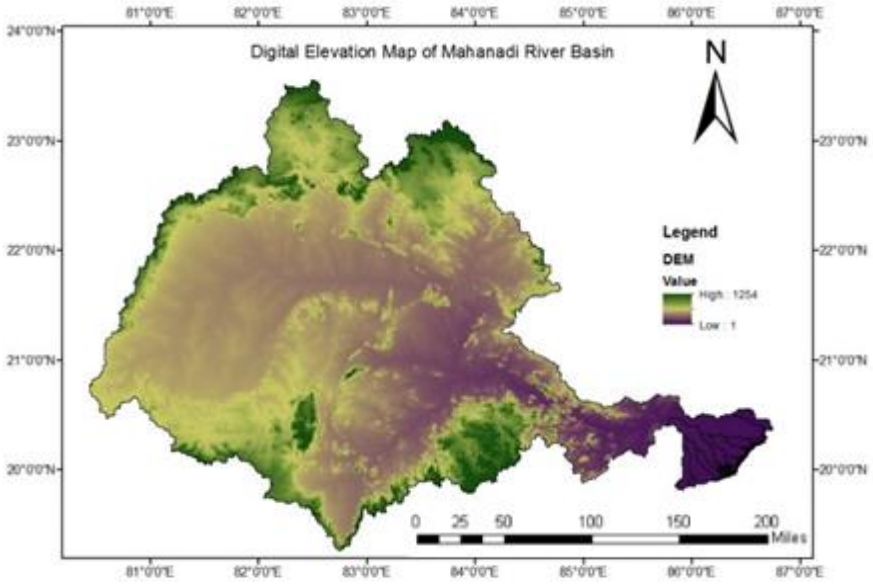


Figure 3.3: Elevation Map of Mahanadi river basin

3.1.3 Climate

Generally the basin experiences four distinct seasons, namely the cold winter, the hot summer, the south-west monsoon and the post monsoon. In the cold winter, the winds are generally light and blow either from the north or the north-east and the atmosphere is bright, thus making winters pleasant. The hot summer commences in March and lasts till the middle of June when the south-west monsoon sets in. Thunderstorms are quite frequent in hot season bringing some rainfall in comparatively higher hilly regions. The highest relative humidity in the basin varies between 68% and 87% and occurs during July/August. The lowest relative humidity occurs during April/May and varies between 9% and 45%. The average highest relative humidity in the basin is 82% and the average lowest relative humidity is 31.6% (Water year Book, CWC, 1997).

3.1.4 Geology

Mahanadi basin predominantly consists of Archaean rocks represented by folded Khondalites, Granite gneisses, and Charnockite. They are inter-banded, and the first two appear to grade into one another. In general downstream part of the river lying in Odisha is dominated by silicate rocks of metamorphic origin.

3.1.5 Rainfall

The catchment of Mahanadi River receives rainfall from south-west monsoon. The average annual rainfall in the basin is 1,463 mm. During the remainder of the year, rainfall is extremely low, rarely exceeding 30 mm per month. The spatial variation in rainfall is moderate in the basin. Average annual rainfall in the most upstream part of the basin is about 1000 mm, increasing toward the central basin part (1300 mm) and further in the most downstream coastal belt of the basin (1700 mm).

3.1.6 Temperature

The coldest and hottest months in the basin are during December ranging from 10° C to 13.7° C and during May ranges from 38°C over the hills to 43° C in the plains respectively. Temperature variation of the basin is from 7°C to 45.5°C. Summer temperatures are averaging around 29ºC and winter temperatures average around 21ºC. In winter, the mean daily minimum temperature varies from 7°C to 12°C. In summer, the mean daily maximum temperature varies from 42°C to 45.5°C.

3.1.7 Soil

Soil has four important properties namely texture, erosion, slope and its productivity respectively. Further texture is classified into four groups such as fine, medium, coarse and rocky. About 41.95 % of the area comes under fine textured soil followed by 51.27 % area under medium textured soil. The Mahanadi Delta is a basin having huge amount of silt deposit that drains a large land mass of the Indian subcontinent into the Bay of Bengal. Primary soils available in the basin are Black soil, Red soil, Yellow soil, Brownish Red to Yellowish Red soil and Dark Gray Coastal alluvial soil as shown in Figure 3.4.

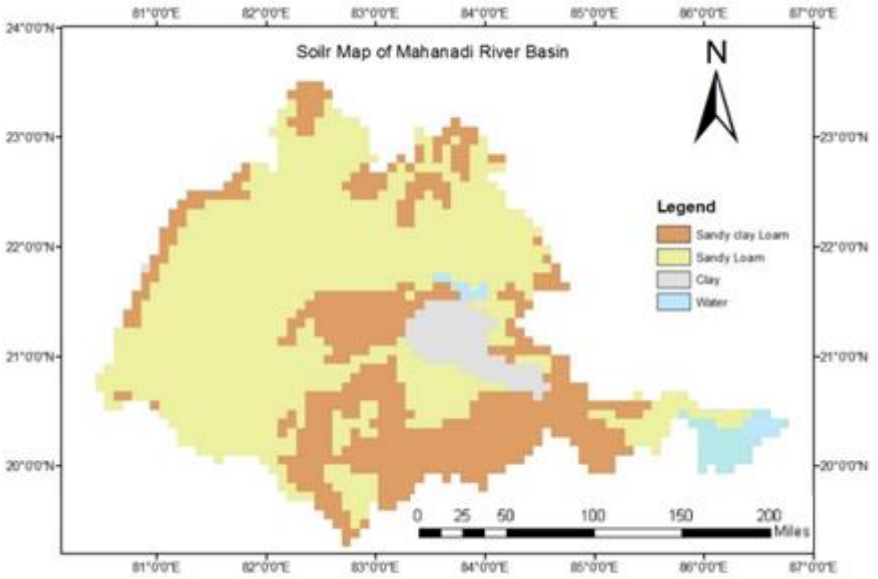


Figure 3.4: Soil Map of Mahanadi river basin

3.1.8 Land use and Land cover

Mahanadi valley is best known for its fertile soil and flourishing agriculture, which primarily depends on a network of canals that arise from the river. Rice, oilseeds and sugarcane are the principal crops cultivated in the Mahanadi valley. The basin has a culturable area of about 79,900 km^2 which is about 57% of the basin area and 4% of the total culturable area of the country. Except in the coastal plains of Odisha, the basin has an extensive area under forests. The sparse vegetation of the highlands contrasts with the moderately luxuriant vegetation of the river valleys. The coastal plains of Odisha, with a high incidence of rainfall, are predominantly rice growing areas. The land utilization pattern of Mahanadi river basin comprises of 37.275% forest area, 10.432% cultivated area, 9.137% area with other uncultivated lands excluding fallow land, 4.967% fallow land and 38.187% net sown area as shown in Figure 3.5. Cultivated area is the total area used for sowing two or more crops in one calendar year. The Net Sown Area is the area sown for each crop but is counted only once. Out of the total annual irrigation water demand of 11km^3 in the basin, the Kharif season utilizes 7km^3 and Rabi season uses 4km^3.Major land use and associated water use changes that have taken place in this basin in the 20th century are related to intensive irrigation of agricultural areas.

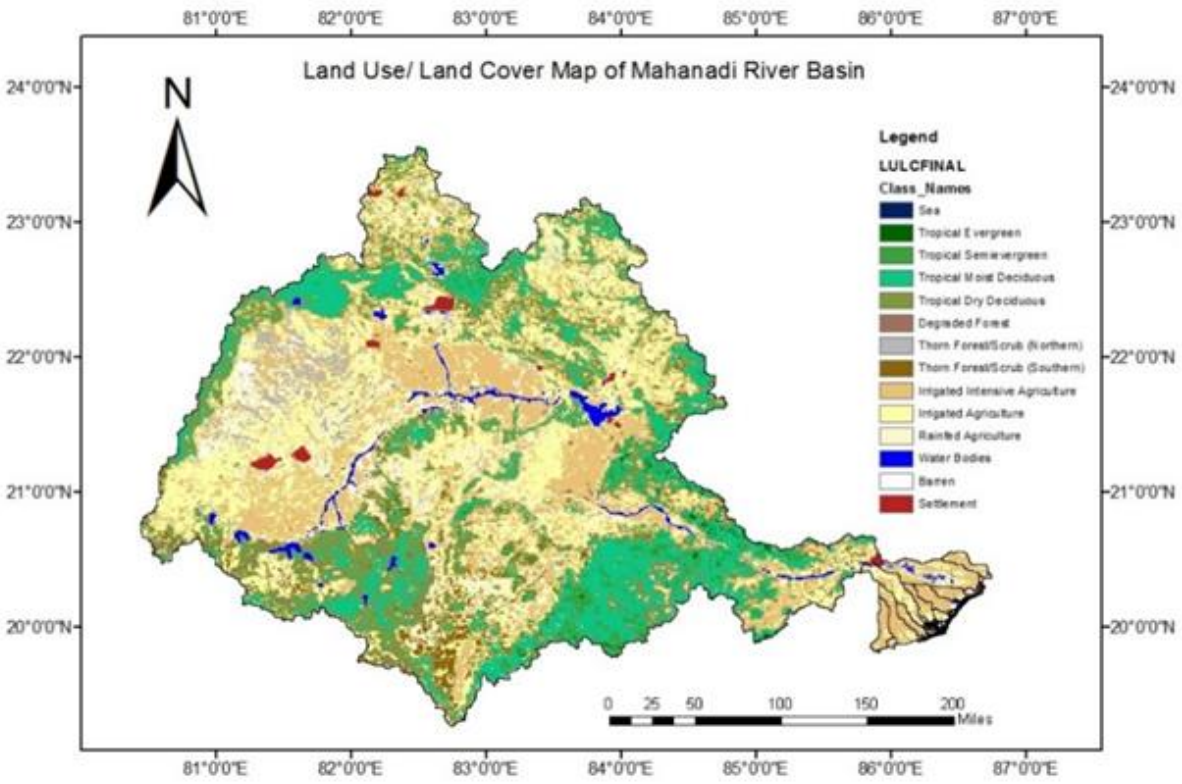


Figure 3.5: Land use land cover Map of Mahanadi river basin

3.1.9 Floods Problem in Mahanadi Basin

The catchment area of Mahanadi is divided into two distinct reaches (i) Upper Mahanadi and (ii) Mahanadi Delta. The upstream catchment of Mahanadi is mountainous and has a steep slope. The catchment lies directly on the south west monsoon track and as such receives heavy rainfall during the summer monsoon. Besides, the catchment area close to the sea is prone to heavy rain brought about by the cyclones generated in the Bay during September-November. Thus, the catchment has the potential of producing a very high flood. The delta area is plain and has a flat slope. Due to flat topography of the delta area the excess flood water is not discharged into the sea quickly and as a consequence, Mahanadi Delta area gets flooded when peak flood discharge exceeds a certain limit. Upper Mahanadi area upstream of Naraj does not have any significant flood problem due to topography except few places in Ib, Bheden and Tel river catchments. The Mahanadi delta covers about 76% of this central area which is having flood prone. The existing embankment system in the Mahanadi delta mitigates floods up to 28,400 m^3/s which is also a 5 year return period flood at Naraj though the escapes provided in the embankments start functioning at different stages of floods starting from 17,000 m^3/s to

25,500 m^3/s (Khatua and Mahakul, 1999). Improvement of structural measures has been advocated to mitigate floods up to 35,000 m^3/s (which is a 10 year return period flood at Naraj). Beyond 35,000 m^3/s structural mitigation measures are found to be economically prohibitive and hence, non-structural mitigation measures are proposed (Khatua and Patra, 2004). In the last decade, high floods exceeding 35,000 m^3/s at Naraj occurred in the years 2001, 2003, 2006, 2008, 2011 and 2013. These floods have caused significant damages to crops, life, and property. The floods of 2008 and 2011 caused a financial loss of more than US $ 400 million (Flood Workshop, 2011). An analysis of historical records of flood events from 1969 to 2011 show that 69% of the major floods are due to the contribution of flow from the middle reaches, 23% due to joint contribution of Hirakud dam and middle reaches while 8% are caused due to contribution from Hirakud releases only (Mishra and Behera, 2009; Parhi et al., 2012). In spite of the presence of Hirakud dam, the Mahanadi delta has been subjected to recent devastating floods. The following are the major reasons reported for the recent high floods in the basin: (i) the Govt. of Odisha states that floods have been occurring due to heavy rainfall in the middle reaches of the basin and accordingly they have proposed a second reservoir (in between Hirakud and Naraj) for controlling the floods (Khatua and Patra, 2004), (ii) some are also of the opinion that releases from the Hirakud dam in late monsoon season might be causing the floods in the delta region as it is regulated by a rule curve which was formulated early in 1988 and that this needs to be modified according to the present rainfall pattern (Flood Workshop, 2011), (iii) also, there are number of irrigation dams in the catchment of Hirakud dam and these belong to the state of Chhattisgarh. During periods of high rainfall in the Hirakud catchment, very high flows are released from these irrigation dams in Chhattisgarh without informing the Hirakud reservoir authorities in Odisha, thus making it difficult for them to control the floods.

3.2 DATA COLLECTION

As the study comprises of flood frequency analysis, development of flood formulae on regional frequency analysis and flood forecasting and inundation mapping, the collection of data varies in a wide range but all are confined to Mahanadi basin only. The basic data required for the study area are daily rainfall, daily discharge data, river cross sections, water level, Soil types, Land-use/ Land-cover, and topographical map. The frequency, length and other details of various data used in the study area are presented in Table 3.1, which were collected and procured from different sources. They were analyzed and transformed for proper use as input to the models.

Table 3.1: List of data and the sources of their collection

S. No.	Station	Data type	Source	Frequency	Period
1	23 G & D sites of Mahanadi Basin	Discharge & Water level	Central Water Commission, Bhubaneswar, and India- WRIS (Water Resources Information System of India)	Daily	1971-2011 (Varying between 10 to 41 years among different G & D sites)
2	13 Rain gauge sites of Mahanadi basin lying in Odisha	Rainfall	Water Resources Department, Govt. of Odisha	Daily	2000-2009
3	34 Districtwise Raingauge sites of Mahanadi basin	Rainfall	India Water Portal	Monthly	1901-2010
4	Mahanadi Basin	SRTM-DEM 90m	CGIAR-CSI (http://srtm.csi.cgiar.org)	-------	-------
5	Mahanadi Basin	Soil Map	NBSS & LUP	------	-------
6	Mahanadi Basin	Land-use/Land-cover	USGS	-------	-------
7	Mahanadi Basin	MODIS Surface Reflectance Data	NASA	Daily	2001, 2003 and 2008
Derived Data					
Sl. No.	**Area**	**Data type**			**Source**

1	Mahanadi Basin	Catchments include area (CA), perimeter (P), mean elevation (E) , average slope (SL), drainage density (DD), longest stream length (LSL), compactness ratio (CR), form factor (FF), shape factor (SF), percentage of basin forested (PBF), percentage of basin agricultural (PBA), percentage of basin barren (PBB), percentage of sandy clay loam (PSCL) and percentage of sandy loam (PSL)	90 m SRTM data
2	Mahanadi Basin	Annual mean precipitation (AMP)	Water Resources Department, Govt. of Odisha

The locations of the rainfall and discharge site over the basin and basin lying in Odisha are shown in Figures 3.6, 3.7, 3.8 and 3.9 respective. For derived data various thematic map such as slope map, flow direction map, flow accumulation map and river network map are required and shown in Figure 3.10. All discharge and rainfall stations name with geographical coordinate is presented in Appendix-1.

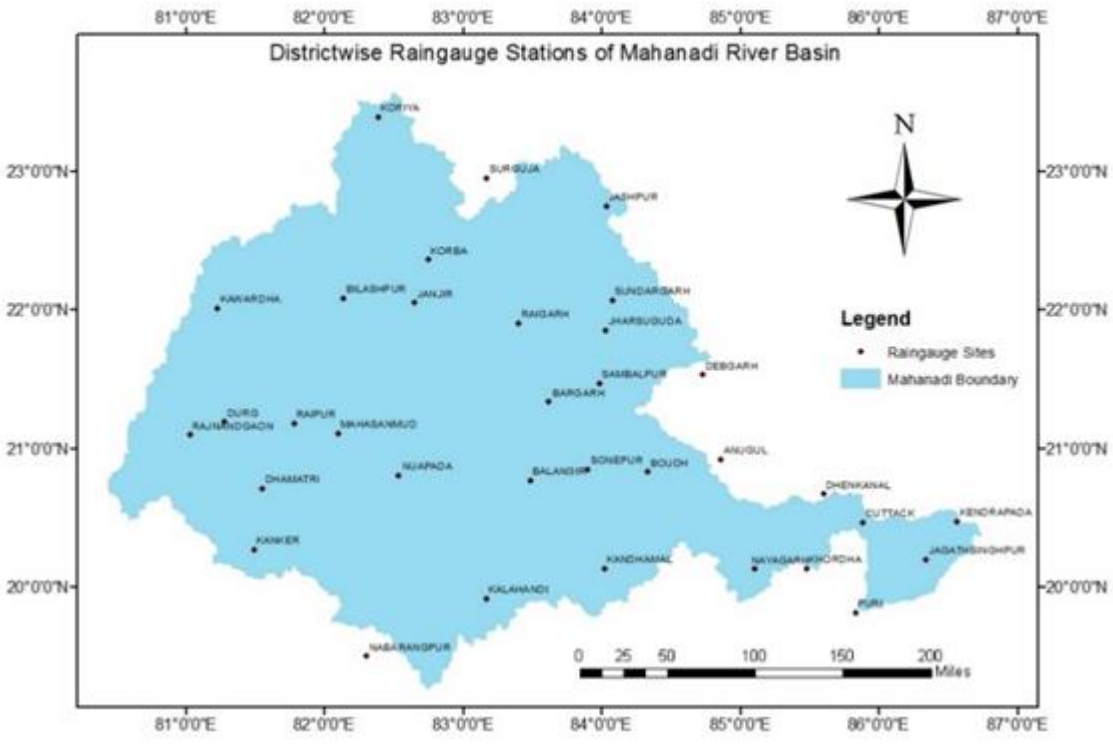


Figure 3.6: Location of Rainfall Stations in Mahanadi River Basin

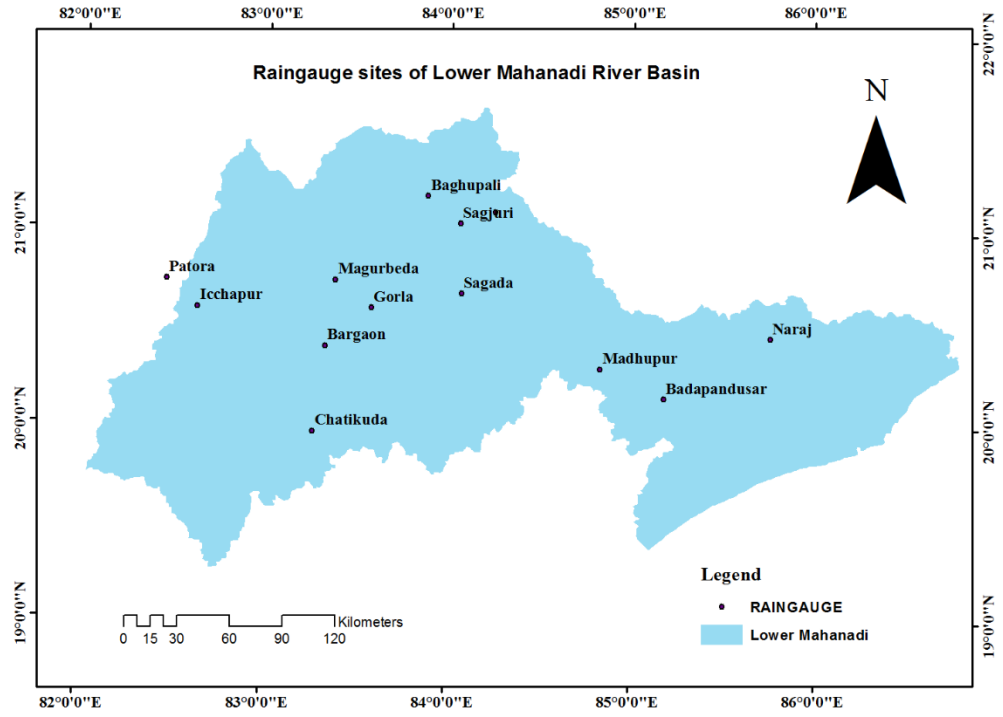


Figure 3.7: Location of Rainfall Stations in Odisha of Mahanadi River Basin

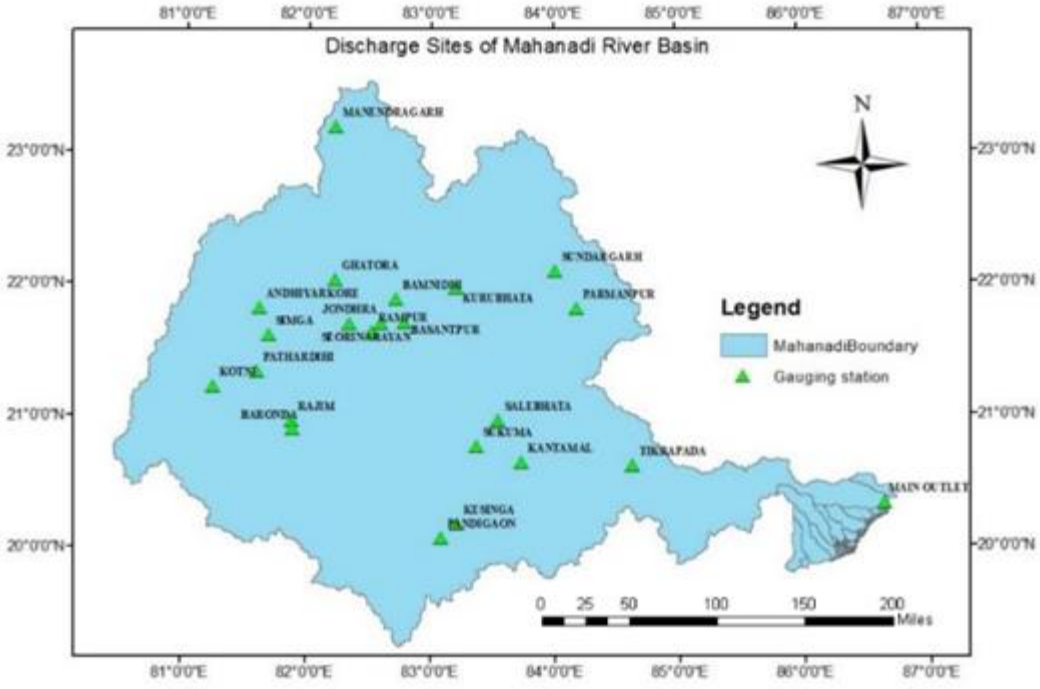


Figure 3.8: Location of Discharge Sites in Odisha of Mahanadi River Basin

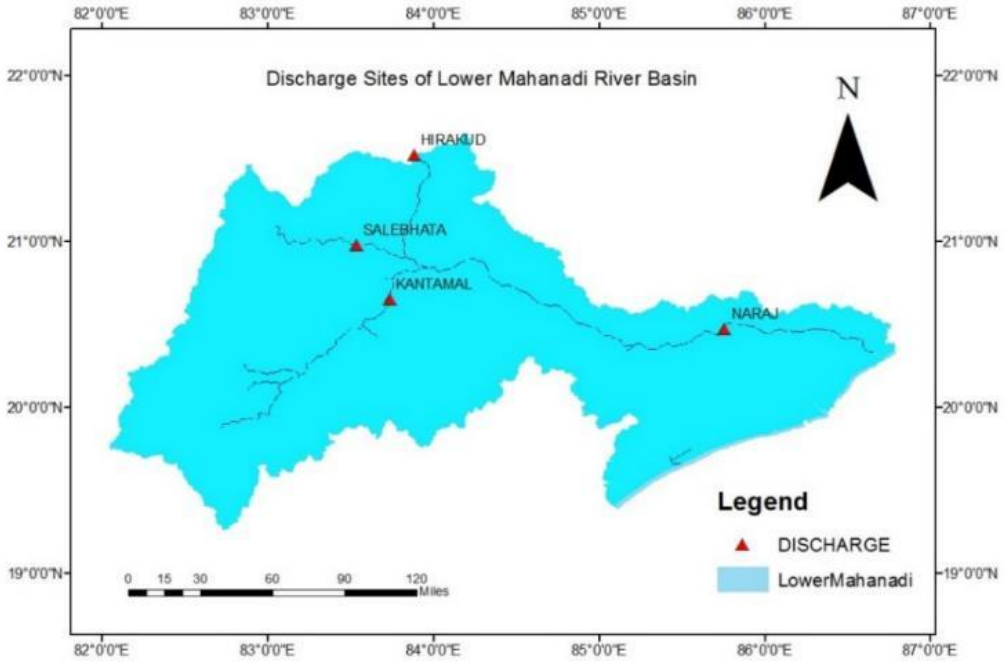
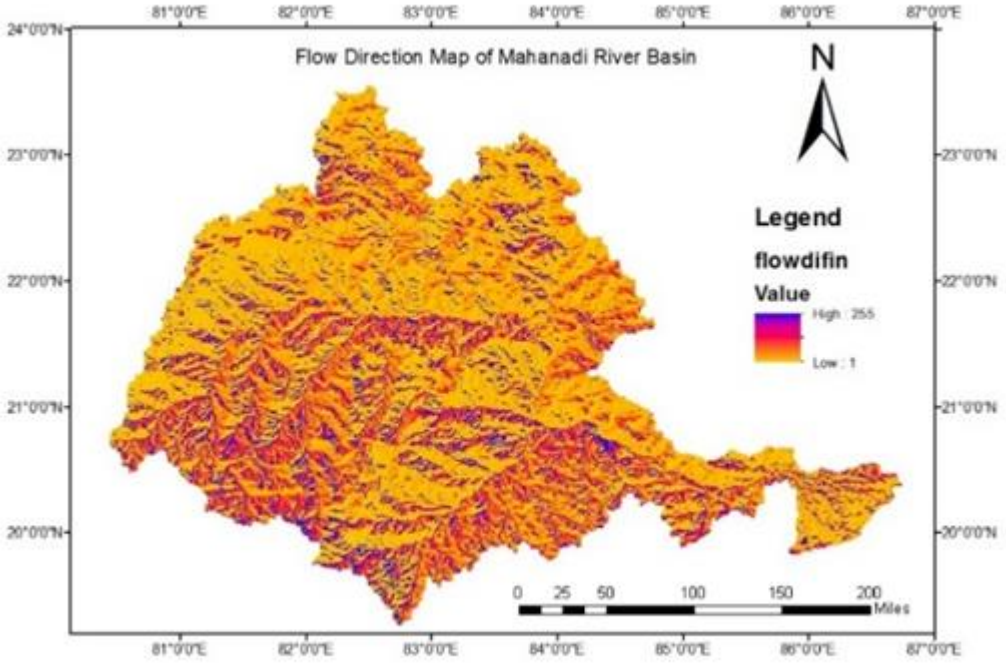
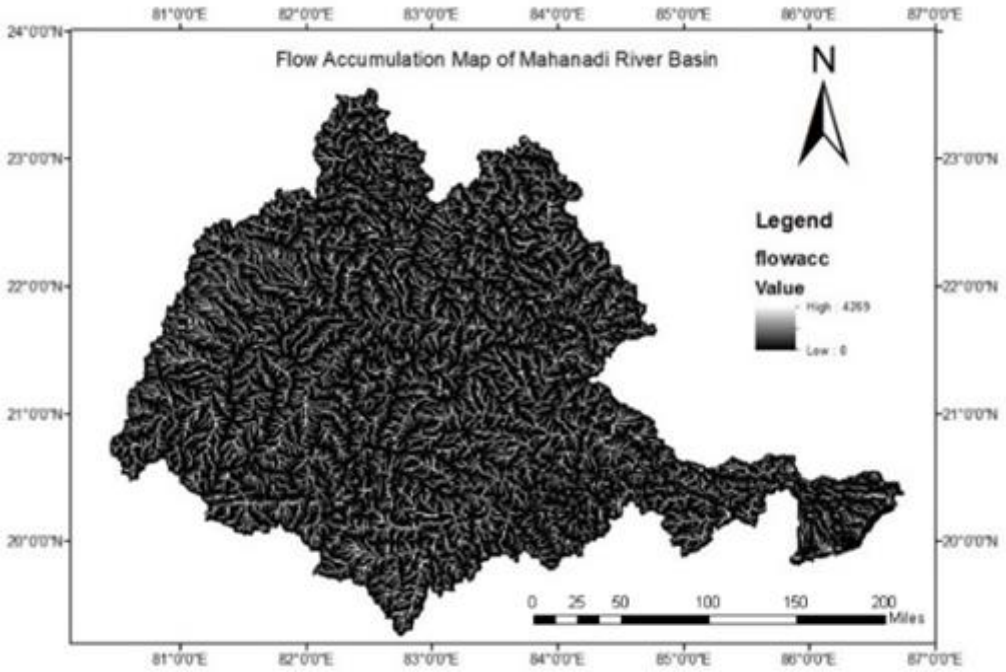
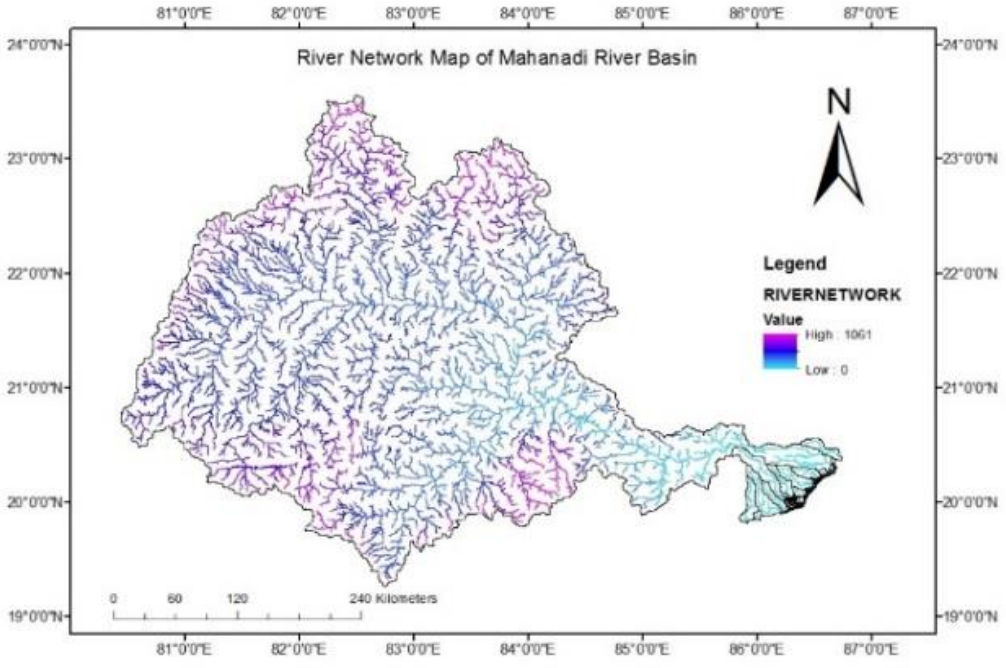


Figure 3.9: Location of Discharge Sites of Lower Mahanadi River Basin




Flow Accumulation Map of Mahanadi River Basin
N
Legend
flowacc
Value
High : 4269
Low : 0
0 25 50 100 150 200
Miles
81°0'0"E
82°0'0"E
83°0'0"E
84°0'0"E
85°0'0"E
86°0'0"E
87°0'0"E
24°0'0"N
23°0'0"N
22°0'0"N
21°0'0"N
20°0'0"N


River Network Map of Mahanadi River Basin
N
Legend
RIVERNETWORK
Value
High : 1061
Low : 0
0 60 120 240 Kilometers
81°0'0"E
82°0'0"E
83°0'0"E
84°0'0"E
85°0'0"E
86°0'0"E
87°0'0"E
24°0'0"N
23°0'0"N
22°0'0"N
21°0'0"N
20°0'0"N
19°0'0"N

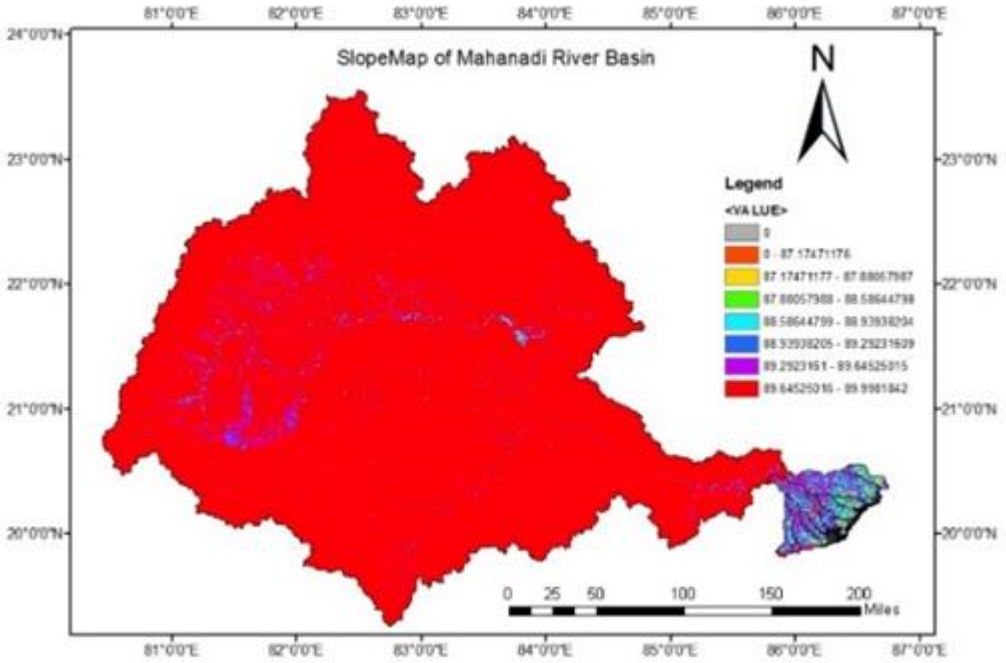


Figure 3.10: Thematic Maps of Mahanadi River Basin

3.3 DATA PROCESSING

A number of processes as listed below were applied to the raw data before they were used for analysis using Excel and Statistical software.

- Screening of data series
- Scrutiny by multiple time series plots
- Checking against the data limits
- Filling of missing values
- Removal of outliers and inconsistencies

Further details of the data used are given in the individual chapters i.e., chapters 4, 5, 6 and 7.

CHAPTER-4
A FRAMEWORK FOR THE SELECTION OF THRESHOLD IN PARTIAL DURATION SERIES MODELING

4.1 INTRODUCTION

In India, floods are very common phenomena causing severe damage to human life, to livestock, to agriculture and to property. Hence, an in-depth understanding of the probabilistic behaviour of such floods is necessary for efficient planning, design and operation of hydraulic structures. Frequency analysis can be used to obtain such knowledge, and develop a relationship between peak values and recurrence interval (Meng et al., 2007). In flood frequency analysis, a series is a convenient sequence of data, such as hourly, daily, seasonal or annual observations of a hydrological variable. Flood frequency analysis can be based on the annual maximum flood (AMF) approach or the partial duration series (PDS) approach, also called peak- over- threshold (POT) approach. An AMF sample is constructed by extracting the maximum value of each year from a daily flow series (e.g.,annual flood), i.e. only one event per year is retained. On the other side, the PDS approach to hydrologic frequency analysis consists in retaining all peak values that "exceed" a certain base level 't' usually called "threshold", thus, there could be more than one extreme events or no events would be selected in any year (Norbiato et al., 2007). PDS approach has lot of advantages, (i) it consists more peak events by selecting the appropriate threshold hence to capture more information regarding the flood phenomena. However, some annual floods may not even be selected as flood events in the PDS approach if their values are less than threshold, (ii) it analyses both, the time of arrival and the magnitude of peaks, (iii) it provides extra flexibility in the demonstration of floods and a complete explanation of the flood generating process. Above these advantages, PDS approach remains under-used and unpopular due to the nonexistence of general framework regarding different approaches,which includes the threshold selection criteria and model hypotheses about the independence, stationarity and distribution of flood peaks. Many researchers proposed methodologies based on the average number of peaks per year for threshold selection for different specific climatic and geographic regions (Taesombut and Yevjevich, 1978; Konecny and Nachtnebel, 1985). Besides, other researchers recommended on the basis of a given return period to choose the threshold level.

Dalrymple (1960) recommended a return period of 1.15 years for threshold selection. Waylen and Woo (1983) and Irvine and Waylen (1986) suggested a return period around 1.2–2 years for Canadian rivers.

The American Society of Civil Engineers (1949) used PDS and recommended the value of λ (i.e., average no. of flood events per year) which does not consider more than 3 to 4 peak flood values in any one year above a threshold discharge. The US Geological Survey recommended that λ should be equal to three.

Cunnane (1973) suggested that for estimation of parameters PDS having λ greater than 1.6 per year are more efficient than AMS and magnitude of peaks are modelled by an exponential distribution also, λ equal to 2 is sufficient for flood quantile estimation. UK Flood Studies Report recommended that λ should be in the range between 3 and 5, (National Environment Research Council, 1975). The best results were obtained by McDermott and Pilgrim (1982) for λ value equal to 1. Few researchers (Tavares and da Silva, 1983; Jayasuriya and Mein, 1985) established that a λ value of 2 or greater is suitable for the compound model following a Poisson distribution of occurrences and an exponential distribution of magnitudes. Hosking and Wallis (1987), studied that the PDS model could improve its performance when λ is equal to 5 in the case of geo-morpho-climatic modeling and λ is equal to 10 in the case of wind velocity modeling.

Rosbjerg and Madsen (1992) recommended a standardized method based on a predefined frequency factor k, and mean and standard deviation of the original series: $t = \bar{x} + k\sigma$ and they set a value of k around 3.

Martins and Stedinger (2001) found that the accuracy of flood quantiles resulting from a GEV-Poisson PDS model is somewhat indifferent to $\lambda \geq 1$. λ values ranging between 3 and 15 for homogeneous Italian regions were estimated by Claps (2003). Rosbjerg and Madsen (2004) conducted a frequency analysis based on rainfall data using a Bayesian framework in Denmark by suggesting a λ value between 2 and 3.

The arrival of the extreme independent events of the PDS model follows the Poissonian assumption, and the magnitude of the exceedances above that threshold is best described by the generalized Pareto (GP) distribution (Begueria, 2005; Trefry et al., 2005; Yuguo et al., 2008).

The popularity of the PDS model for quantifying extreme events is low among experts, as it is often associated with various difficulties. Furthermore, there is no general recommendation existing for choosing an appropriate threshold and thus resulting in difficulties in modeling the PDS coupled with the selection of the appropriate distribution which play a significant role in satisfying the assumptions of the PDS model, and to confirm reliable quantile estimates.

Henceforth, the following observations were made from the previous studies: (a) very little work has been done in the past to evolve any flood frequency distribution by considering peak discharge floods of longer duration using PDS beyond a threshold, as most of the studies are based on extreme rainfall (b) actual observations (peak discharge along with Highest Flood Level using Rating Curve at each location in the field are not considered in most of the cases while selecting the most appropriate threshold values (t) or average peak per year (λ) values. Keeping this in view, the objective of the present research includes formulating the operational guidelines for selection of threshold value based on different concepts and then test the performance of GP/PDS modeling based on the suitability of the GP distribution parameters coupled with the appropriate threshold values. In the present study, extensive field data such as cross-sectional data with water level including highest flood level, discharge data for 22 sampling locations were collected for a period of 1971-2012 from various central and state government agencies. Figures 4.1, 4.2 and 4.3, illustrate discharge, cross-sections with water level and rating curves at two representative sites Tikarapara and Rajim in the Mahanadi river system. The HFLs at Tikarapara and Rajim stations along Mahanadi river are 68.86 m and 279.72 m respectively.

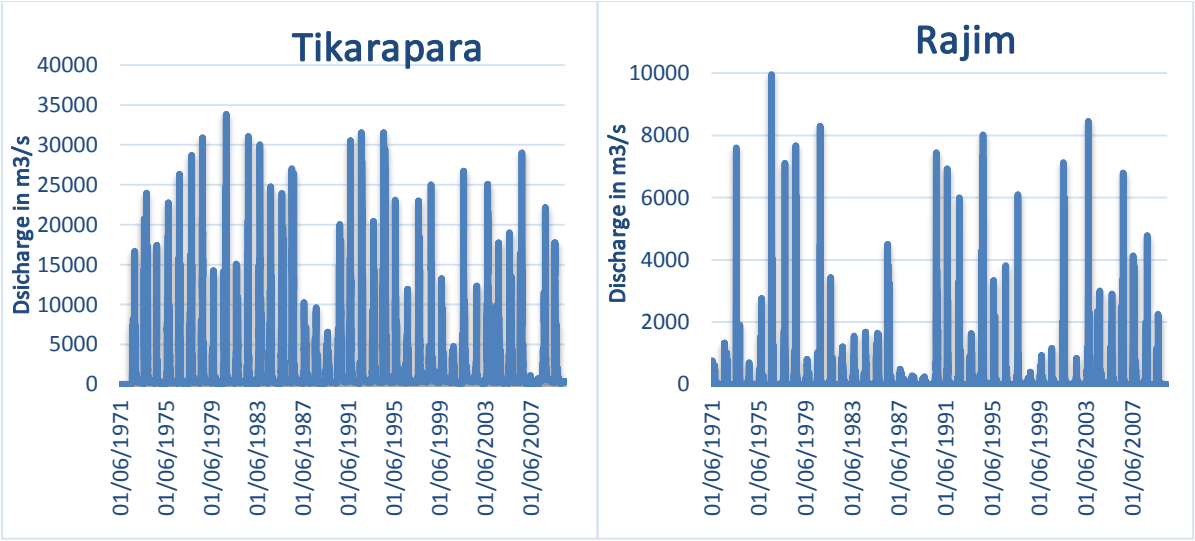
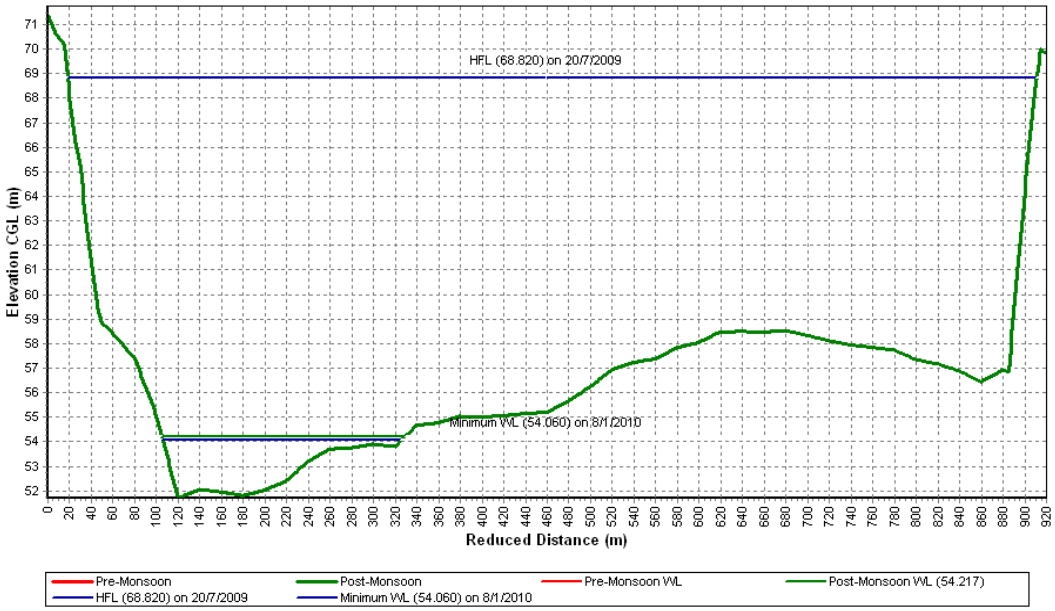


Figure 4.1: Observed discharge data at two sampling locations



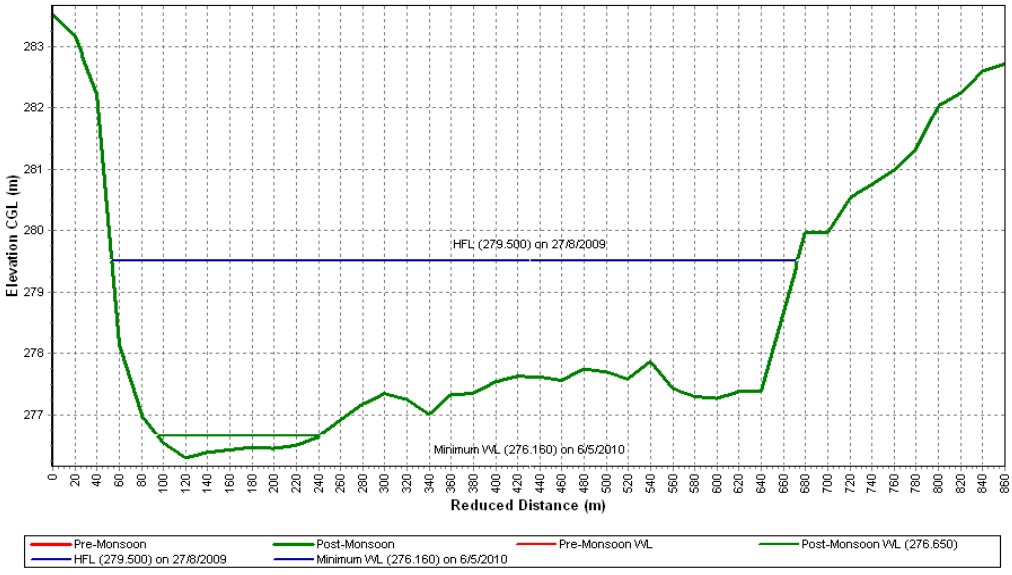


Figure 4.2: Observed water level and cross-sections at two sampling locations

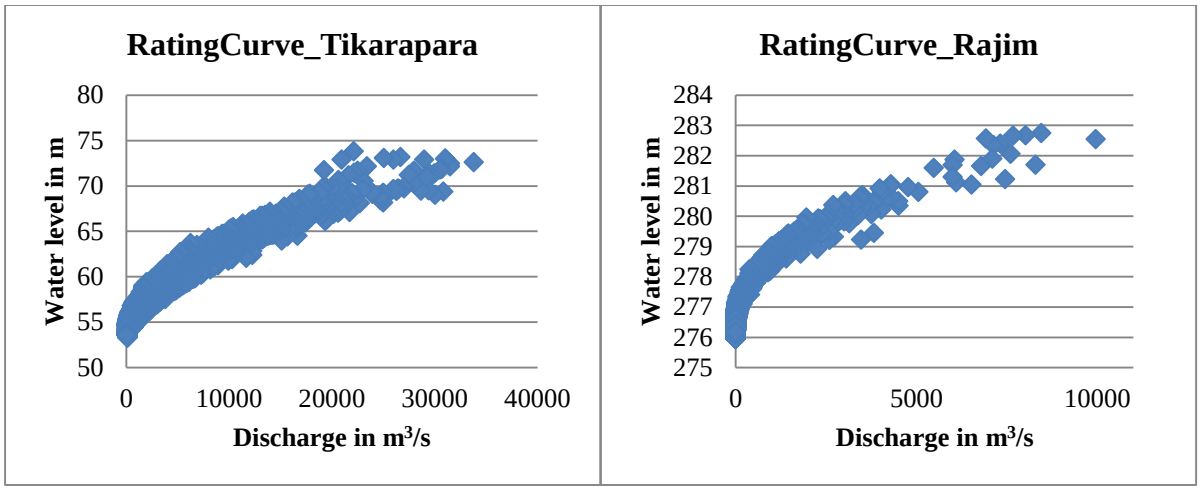


Figure 4.3: Rating curve at two sampling locations of Mahanadi river system

4.2 METHODOLOGY

4.2.1 Overview of Partial Duration Series

In general, the partial duration series is characterised by the average number of peaks each year (λ) or a threshold value (t). Consider a daily time series data, $X = \{X_1, X_2, .. X_i .., X_n\}$ over 'n' years, then the PDS having 'p' values represented as , $M = \{M_1, M_2, .. M_j .. , M_p\}$, which exceed a preferred threshold t, $M_i > t$ with t = constant. Here, the total number of peaks (p) are obtained from the average number of peaks per year, $\lambda = p/n$ and assumed to be identically an independent variable and it follows Poisson distribution.

Generally, the average number of peaks per year (λ) is getting increased, with a decrease of the threshold value (t). Thus, an appropriate threshold value should be selected for the flood frequency analysis. If we set a higher threshold value, it could be independent of the Poisson distribution. However, this can lead to a significant loss of information of high magnitude peaks that cause flooding with increasing uncertainty. In contrast, if more peaks are selected with series dependence of both time interval and magnitudes, it violates the assumption of independence but provides the reliable distribution parameter estimation (Ashkar et al., 1983; Buishand, 1989; Cunnane, 1979, 1985; Langa et al., 1999; Onoz and Bayazit, 2001; Trefry et al., 2005).

4.2.2 Framework to Select the Threshold Values using different Concepts

A Partial Duration Series (PDS) is obtained from the daily streamflow data at all stations by setting the different threshold value acquired from different concepts. They are used by:

(i) Considering an average number of flood peaks per year (λ=1, 1.5, 2, and so on). For λ = 1, the resulting series consists 'n' highest peaks within the 'n' years of data. There are possibilities of having two or more number of peak floods from one year and no flows from the other years as they are lying below the threshold value. The independence condition is one of the hypotheses of the Poisson process. The choice of the threshold should ensure that peak values meet this condition. A considerably low threshold leads to the selection of minor peaks that are not always independent. Raising the threshold should help avoid this situation.

(ii) Considering mean exceedance above threshold ($(\bar{M}_t - t)$ is a linear function of the threshold value t, where $\bar{M}_t$the average value of exceedances and this criteria is equivalent for choosing the threshold to maximize the stability of the POT distribution parameter estimates.

(iii) Considering observed Highest Flood Level (HFL) at each sampling location and the corresponding discharge values using Rating Curve. The observed HFL helps to calculate the stage (WL) and based on that stage information we get the threshold discharge value by plotting the stage-discharge curve. The values of discharge for stage above HFL cause real flood situation in the field and causes flood inundation at various locations in Mahanadi River Basin. In fact, in most of the previous studies the actual observations (peak discharge along with Highest Flood Level) in the field are not considered while selecting the threshold values and average peak per year (λ) values.

However, threshold selection is strongly related to the hypothesis of independence, stationarity and the combined GP-Poisson assumption. Hence after selection of threshold from different concepts the above tests must be analyzed. Although the GP distribution is considered as the best distribution for the PDS series, studies on their suitability is required. Hence two measures such as the L-moment ratio diagram and the stability of the GP parameters are used.

4.2.2.1 Randomness Test for PDS

Independence criteria are essential to any statistical frequency analysis.

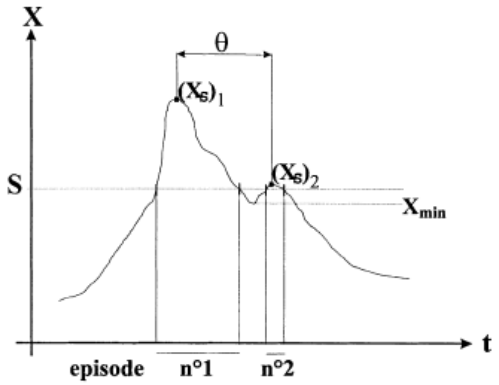


Figure 4.4: Inter-Flood Duration Criteria (Source: Lang et al., 1999)

The Water Resources Council (USWRC, 1976) suggested that successive flood events be separated by at least as many days as five plus the natural logarithm of square miles of the basin area. In addition to the random condition that the intermediate flows between two consecutive peaks must drop below 75% of the lowest of these two flood events. According to Lang et al., (1999), if X is a random variable, they define X_S as the maximum value of X in an episode. An episode is defined as a function of a threshold level S: it begins when X(t) exceeds S and ends when X(t) falls below the level S. Referring to Figure 4.4, this condition means that the second flood peak $(X_S)_2$ must be rejected if:

$$\theta < 5\ days + log(A) \quad or \quad X_{min} > \left(\frac{3}{4}\right) min[Q_1, Q_2] \tag{4.1}$$

where A is the basin area (miles^2) and Q_i is the maximum daily discharge of flood number i. Cunnane (1979) provides the following criteria. The second flood peak $(X_S)_2$ must be rejected if:

$$\theta < 3T_p \ \ or \ \ X_{min} > \left(\frac{2}{3}\right)(X_S)_1 \tag{4.2}$$

where T_p is the average time to peak.

The autocorrelation test is another convenient test to check the independence criteria in the PDS (Miquel, 1984) which comprises the number of peaks selected irrespective of their occurrence (Gordon et al., 2004).

The randomness test (auto-correlation test) is done to find independent flood peaks from all the data sets above given threshold values at each station. When these datasets are independent, the autocorrelation function for all lags other than zero is observed to be zero. The correlation coefficient (r_k) is calculated using the equation:

$$r_k = \frac{\sum_{i=1}^{N-k}\left(x_i - \bar{x}\right)\left(x_{i+k} - \bar{x}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \tag{4.3}$$

where $\bar{x} = \sum_{i=1}^{N} x_i$ is the overall mean.

Once the data are found to be random in nature, it may be used for frequency analysis.

4.2.2.2 Stationarity Test for PDS

Stationarity is another requirement to detect the presence of trends by means of Mann–Kendall test in the extracted PDS sets. This Mann–Kendall statistic is the most common nonparametric trend test that has no restraints on the trend characteristics. It tests a null hypothesis H_0 (no trend exists in the data) against an alternative hypothesis H_A (existence of increasing or decreasing trend). The null hypothesis H_0 is rejected when the value of the following statistic $Z = |S|/\sigma^{0.5}$ is greater than the $Z_{\alpha/2}$ value with a confidence level $\alpha = 0.05$.

4.2.3 Outline of GP-Poisson Distribution Model

4.2.3.1 Poisson Distribution Model

The occurrence of events exceeding the given threshold (t) is best described by the Poisson distribution which follows a random process. The probability density function in the case of Poisson distribution is given by

$$f(m,\lambda_k) = \frac{e^{-\lambda_k}(\lambda_k)^m}{m!} \tag{4.4}$$

where m is the occurrences of numbers of peaks within the time interval (0,k), and λ is the mean arrival rate (or the average number of peaks per year) which can be calculated from the sample. The Poisson assumption implies that the occurrences of the events are independent in nature.

The Dispersion measure is commonly used for testing the suitability of the fitted Poisson distribution. To obtain the optimum value of λ and to test the adequacy of the fitted Poisson distribution, the Fisher dispersion index (DI) test was done using the following equation:

$$DI = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{\bar{x}} \tag{4.5}$$

where $\bar{x}$ stands for the sample mean.

The Dispersion Index technique is approximately distributed as a Chi-square (χ^2) statistics with (n-1) degrees of freedom for larger values of n. The probability (p) of the hypothesis H_0 is accepted when $0.05<p<0.95$ and H_0 is rejected when $p<0.05$ and the corresponding DI value is equal to 1.

4.2.3.2 GP Distribution Model

The PDS comprises all mutually independent peaks which follow the GP distribution. The GP distribution is defined by a location (μ), scale(σ), and shape (ξ). The Generalized Pareto (GP) distribution describes the highest values from continuous hydrological time series and has been carefully chosen by many authors for Poisson distribution modeling (Van Montfort and Witter, 1986; Hosking and Wallis, 1987; Wang, 1996; Madsen and Rosbjerg, 1997; Martins and Stedinger, 2001). The cumulative distribution function of the GP distribution is given by:

$$P\left(X \le x \middle| \mu,\sigma,\xi\right) = 1 - \left(1 - \xi \cdot \frac{\left(x-\mu\right)}{\sigma}\right)^{\frac{1}{\xi}}, \xi \ne 0 \tag{4.6}$$

$$P\left(X \le x \middle| \mu,\sigma,\xi\right) = 1 - \exp\left(-\frac{x-\mu}{\sigma}\right), \xi = 0$$

For the shape parameter ξ=0, the GP distribution transforms into a simple exponential distribution. Even though the GP distribution is considered as the best distribution for the PDS analysis of their suitability is required. Hence, we used two measures such as the L- moment ratio diagram and the stability of the GP parameters.

L-moment ratio diagram is drawn for each distribution, which shows the theoretical relationship between L-skewness ($\tau_3 = LC_S = \frac{\beta_3}{\beta_2}$) and L- kurtosis ($\tau_4 = LC_K = \frac{\beta_4}{\beta_2}$). The curve representing the L-moment ratios of the GP distribution is given below:

$$\tau_4 = \frac{\tau_3\left(1+5\cdot\tau_3\right)}{5+\tau_3} \tag{4.7}$$

To satisfy the assumption we fit another four additional probability distribution such as generalized logistic (GL), the generalized extreme value (GEV), the log-normal (LN3), and the Pearson Type III (PEIII). When the τ_3 and τ_4 sample estimates are plotted on the L-moment ratio diagram, the dispersion or clustering of the samples points around the theoretical relationship of a particular flood frequency distribution helps in deciding the fittingness of any distribution (Hosking and Wallis, 1993).

The L-moments values are calculated using the following equations:

$$\lambda_1 = \frac{1}{n}\sum_{j=1}^{n} X_j \tag{4.8}$$

$$\lambda_2 = 2\sum_{j=1}^{n-1}\frac{(n-j)X_j}{n(n-1)} - \frac{1}{n}\sum_{j=1}^{n} X_j \tag{4.9}$$

$$\lambda_3 = 6\sum_{j=1}^{n-2}\frac{(n-j)(n-j-1)X_j}{n(n-1)(n-2)} - 6\sum_{j=1}^{n-1}\frac{(n-j)X_j}{n(n-1)} + \frac{1}{n}\sum_{j=1}^{n} X_j \tag{4.10}$$

$$\lambda_4 = 20\sum_{j=1}^{n-3}\frac{(n-j)(n-j-1)(n-j-2)X_j}{n(n-1)(n-2)(n-3)} - 30\sum_{j=1}^{n-2}\frac{(n-j)(n-j-1)X_j}{n(n-1)(n-2)} + 12\sum_{j=1}^{n-1}\frac{(n-j)X_j}{n(n-1)} + \frac{1}{n}\sum_{j=1}^{n} X_j \tag{4.11}$$

The stability of the GP parameter values has a great impact on the quantile estimation and reduces the uncertainty of the change in the threshold value (Begueria, 2005; Deidda et al.,2009). The parameter location (μ), scale(σ), and shape (ξ) are obtained using L-moment (L-MOM) according to the equations given by Hosking and Wallis (1997):

$$\mu = \lambda_1 - (2+\xi)\lambda_2 \tag{4.12}$$

$$\sigma = (1+\xi)(2+\xi)\lambda_2 \tag{4.13}$$

$$\xi = \frac{1-3LCs}{1+3LCs} \tag{4.14}$$

where L-Cs are the L-Moment coefficients of skewness. The values λ_1 and λ_2 are L-Moment mean (1st order) and L-Moment standard deviations (2nd order) respectively.

Since in PDS modeling the threshold value is known, hence no need to calculate the location parameter. After that, a quantile (QQ) plot shows the relation between the observed return levels against the estimated return levels associated with the ith Gringorten plotting position, $F_i = \frac{i-0.44}{n+0.12}$, where n is the number of data, and i is the rank of the data. Then, the performance is evaluated using relative root mean square error (RRMSE) and relative bias (RBIAS) by the change in the λ value of GP distribution.

The RRMSE and RBIAS are given by:

$$RRMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\frac{Q_E - Q_O}{Q_O}\right)^2} \tag{4.15}$$

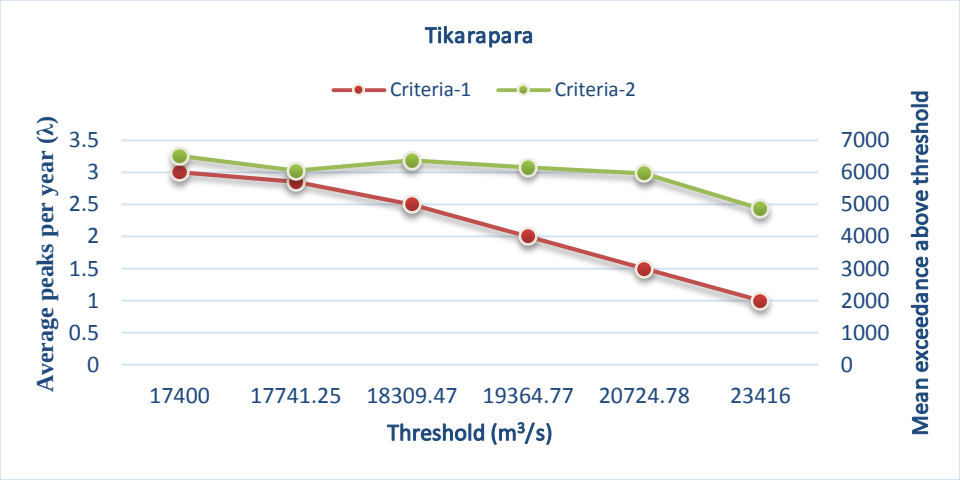
$$RBIAS = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{Q_E - Q_O}{Q_E}\right) \tag{4.16}$$

where, M is the total number of samples, Q_E is estimated quantile of the m^{th} sample and Q_O is the observed data. The smallest values of RRMSE and RBIAS correspond to the best quantile with different λ values for a number of recurrence intervals.

4.3 RESULTS AND DISCUSSION

4.3.1 Threshold Values using different Concepts

Considering the (i) average number of flood peaks per year (λ=1, 1.5, 2, and so on), (ii) mean exceedance above the threshold, and (iii) observed Highest Flood Level (HFL) at each sampling location and corresponding discharge values using Rating Curve are obtained in each case. Results of the above concept dealing with the choice of threshold values are presented in Figures 4.5 for two representative sampling locations of Mahanadi river basin. A threshold Value of 18309.47 m^3/s at Tikarapara station, leading to an average peaks of 2.5 per year and 2256.82 m^3/s at Rajim having average peaks of 2.12 per year, obtained from HFL, are selected. Also, the mean exceedance above the threshold is found to be a stable function of the above threshold. The above analysis is done for all the 22 stations and found that the threshold level having average peaks per year (λ) value from 2 to 3 based on above criteria. The results are shown in Appendix I.



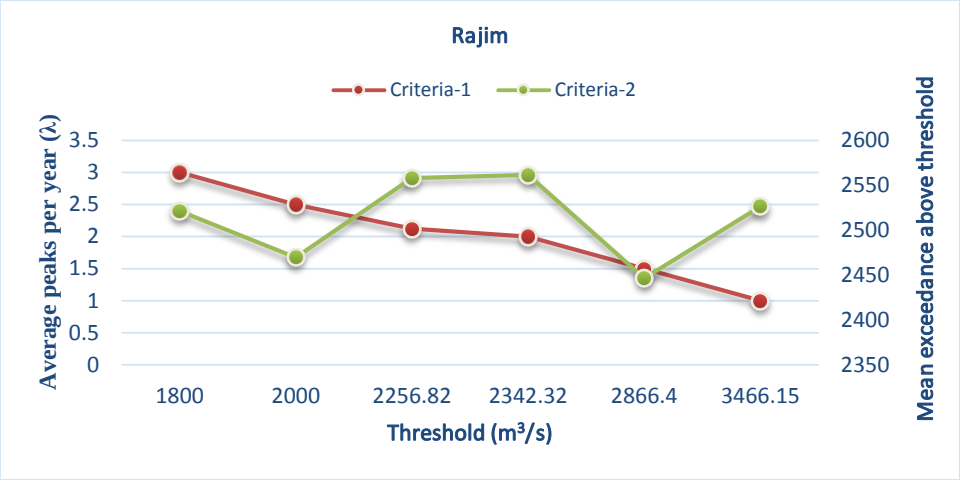
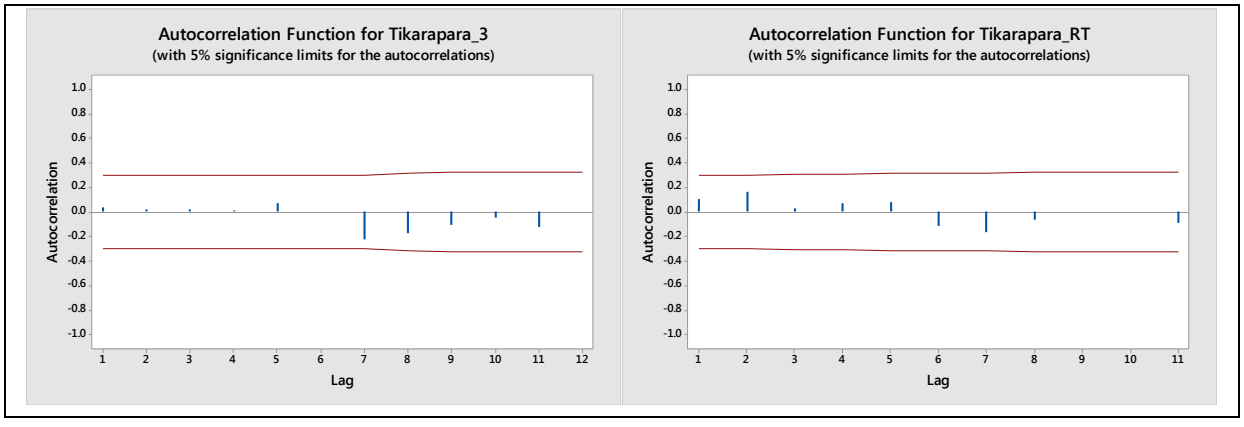


Figure 4.5: Threshold values obtained using different concepts

After selection of threshold values different analysis such as the hypothesis of independence, stationarity and the combined GP-Poisson assumption were tested.

4.3.2 Randomness Test for PDS

Following the concept discussed earlier, a minimum interval of 7 to 8 days between events for all the 22 discharge sites are used to ensure that data used in the further analysis are independent. The randomness test is also done for all the sites to find independent flood peaks from all the data sets above given threshold values. The auto-correlation function (ACF) of the original discharge data and extracted PDS data are tested. The ACF plot indicates that the partial duration series of flood data are independent and can be used for further analysis. Figure 4.6 illustrates the ACF plots for two representative stations Tikarapara and Rajim, for λ=3 and HFL obtained from rating curve and remaining are shown in Appendix I.



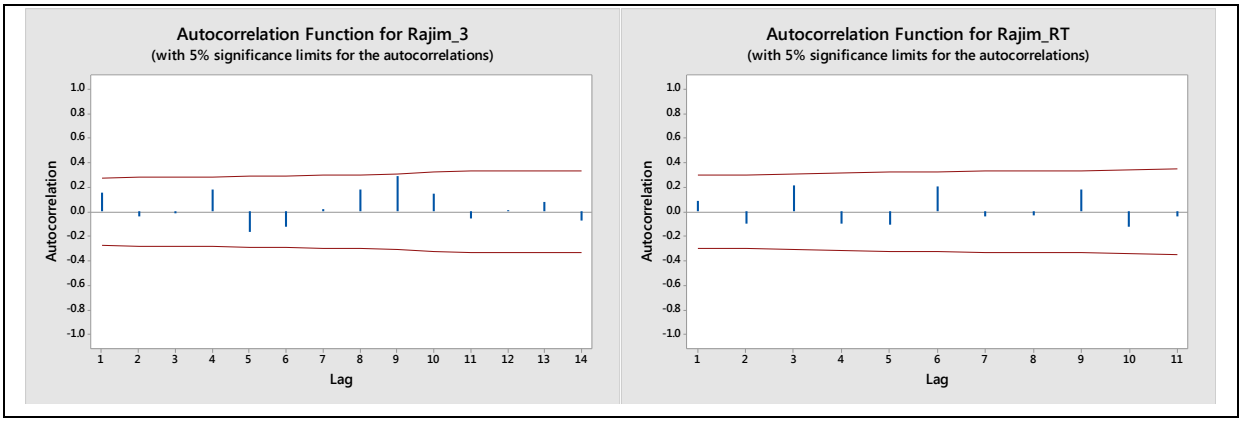


Figure 4.6: Autocorrelation function of the discharge data of Tikarapara and Rajim

4.3.3 Stationarity Test for PDS

From the Mann–Kendall trend test, probabilities are computed from absolute Z Statistics and test the null hypothesis H_0 (no trend exists in the data) against an alternative hypothesis H_A (existence of increasing or decreasing trend) with a confidence level $\alpha = 0.05$ and found no trend in the PDS at almost all stations. At stations Paramanpur and Simga, the results are showing a decreasing trend for $\lambda=1$ and at Manendragarh the results are showing decreasing trend for almost all threshold values.

4.3.4 Verification of Poisson distribution with change in Threshold

The arrival of independent PDS peaks follows the Poisson distribution. To verify the above assumption we have calculated the dispersion index (DI) and test the probability (p) of the hypothesis for all 22 stations. Figure 4.7 shows a box plot of the DI values obtained at the different stations with change in the threshold values. The results indicates increase in the DI values with an increase in the threshold and satisfies the probability (p) hypothesis that the data follows a Poisson process and is accepted (within the limits of $0.05<p<0.95$).

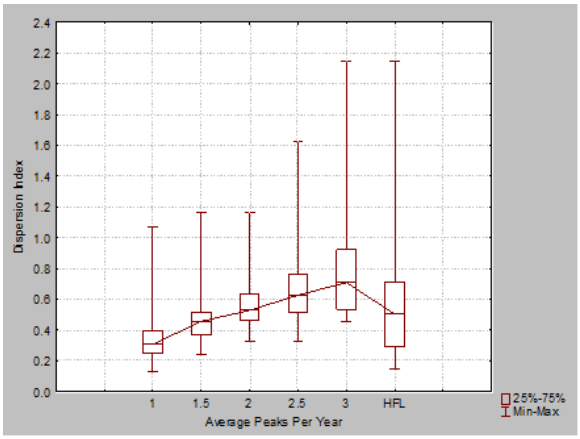
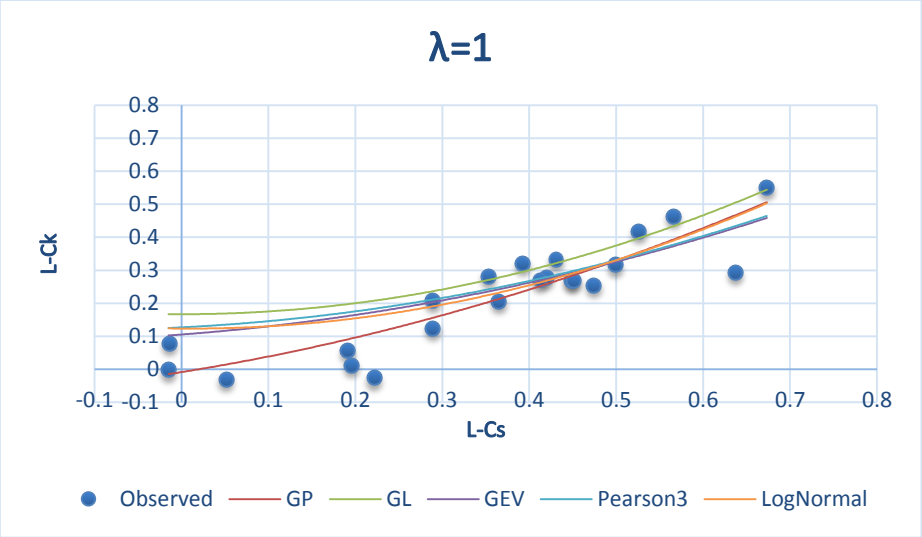
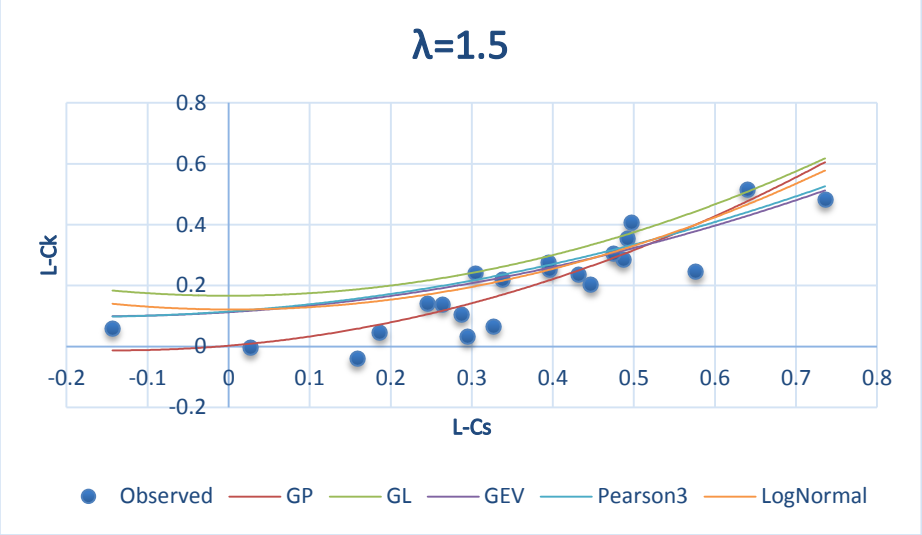


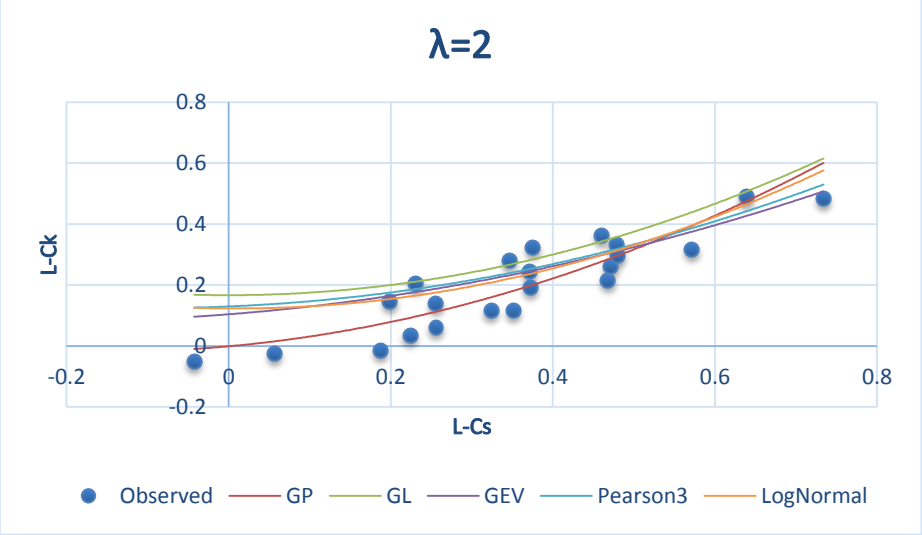
Figure 4.7: Poisson Distribution for Peak Occurrences

4.2.3.2 GP Distribution Model

According to the assumption, the GP distribution is considered as the best distribution for the PDS series. To verify the above assumption a study towards their suitability is required, hence we used two measures such as the L- moment ratio diagram and the stability of the GP parameters. L-moment ratio diagram shows the clustering of the sample datasets around the theoretical relationships between L-Cs and L-Ck of different probability distributions. The suitability of the GP distribution (using combination of equations 4.6-4.11) along with other four flood frequency distributions [viz,Generalised Extreme Value (GEV), Generalized Pareto (GP), Pearson Type 3, Generalized logistic (GL) and Lognormal] describing different threshold values are shown in Figure 4.8. It is observed that in the case of λ equal to 1 (higher threshold value) the sample L-Cs and L-Ck values distribute over a widespread region signifying that variety of distributions can be fitted for frequency analysis. When the λ values increases, the sample L-Cs and L-Ck values tend to group around the theoretical L-Cs and L-Ck curve of the GP distribution in almost all stations and found to be well represented in the cases for equal and more than λ=2. However, the stability occurs as we increase the λ value beyond 2. Interestingly, the flood inundation and damage takes place for the λ value ranging between 2 and 3. Also, the HFL of almost all stations varying within this range.


λ=1
L-Ck
L-Cs
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Observed
GP
GL
GEV
Pearson3
LogNormal


λ=1.5
L-Ck
L-Cs
0.8
0.6
0.4
0.2
0
-0.2
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Observed
GP
GL
GEV
Pearson3
LogNormal


λ=2
L-Ck
L-Cs
0.8
0.6
0.4
0.2
0
-0.2
-0.2
0
0.2
0.4
0.6
0.8
Observed
GP
GL
GEV
Pearson3
LogNormal

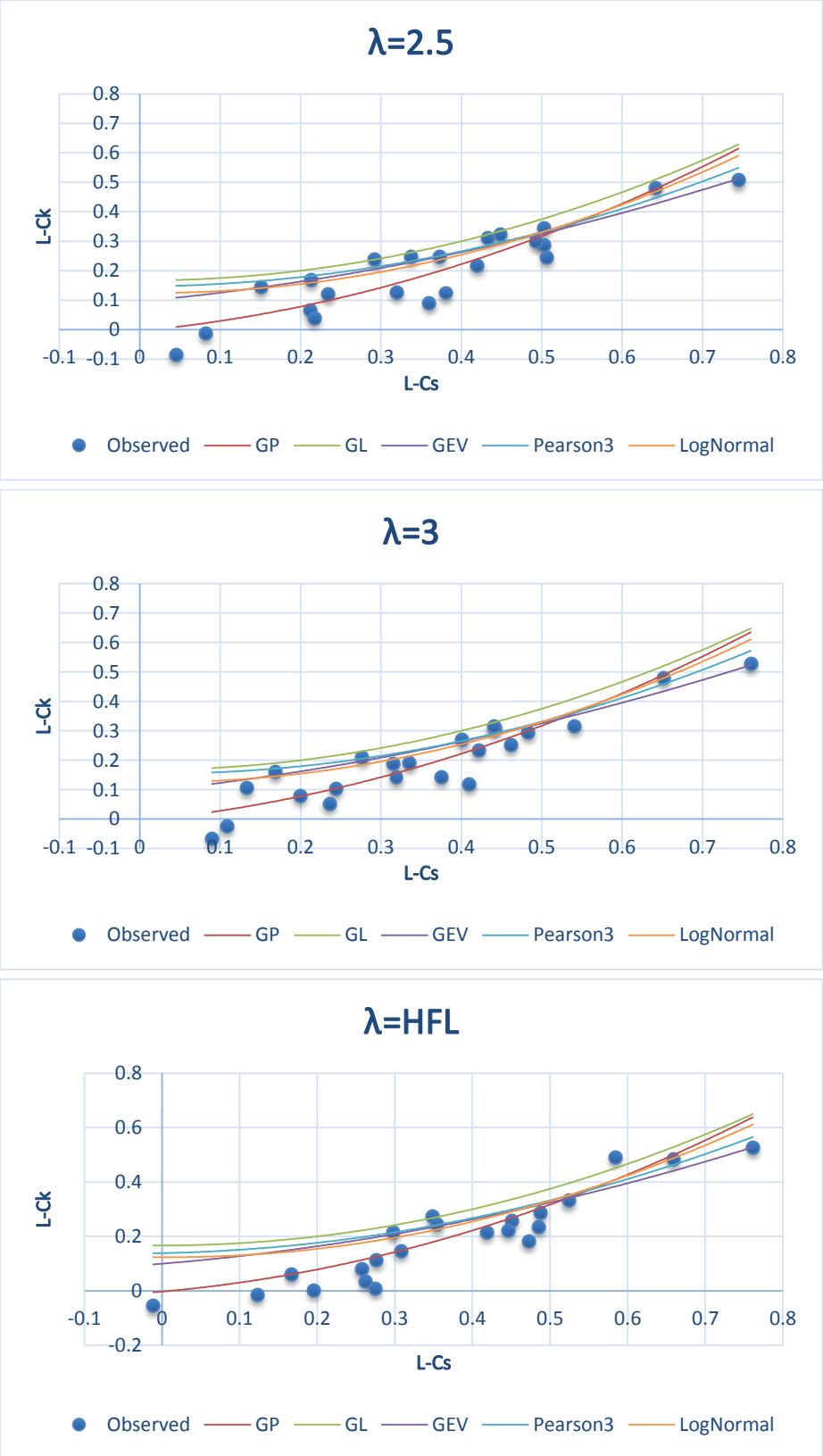
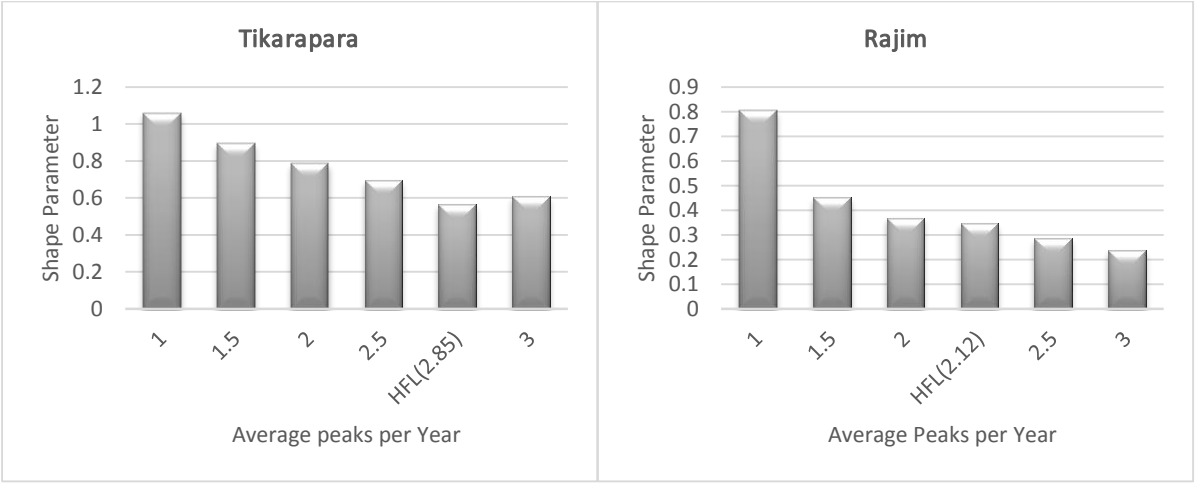


Figure 4.8: L-moment diagram for the suitability of PDS

From Figure 4.8, it has been observed that the GP distribution provides best results in comparison to other frequency analysis in all the cases. Now, it was essential to evaluate the change in the GP distribution parameters, namely scale(σ), and shape (ξ) for different threshold (t) or average peaks per year (λ) values to obtain the optimum threshold value having constancy in model parameters. It has been observed that there is a gradual decrease in the value of scale(σ), and shape (ξ) parameters at all stations when the threshold values decreases and, PDS peak flood data increases. As a result, the parameters of the GP distribution performed excellently to identify the optimal number of exceedance based on threshold value. Figure 4.9 illustrates that the steadiness of the parameters with the increase in λ values for two representative sampling locations Tikarapara and Rajim. From the Figure, it is noted that the shape (ξ) and scale(σ) parameters of Tikarapara station begins to stabilise at threshold (17741.25 m^3/s) obtained from HFL and corresponding discharge using Rating Curve at λ=2.85. Similarly at location Rajim, the shape (ξ) and scale(σ) parameters begins to stabilise at λ=2 and more stabilization occurs by increasing threshold values beyond λ=2. It has been observed that in all the 22 sampling locations, the beginning of stabilization of shape parameters takes place for λ values ranging between 2 and 3 presented in Appendix 1. In fact, the flood inundation and flood damage also takes place for the higher peaks floods above threshold in the basin. Higher values of λ increases the number of peak discharge values in the PDS and stabilizes the GP distribution parameter. However, it is not causing flood inundation and damage in actual field locations of Mahanadi river basin, India presently and may be used to develop future scenario.



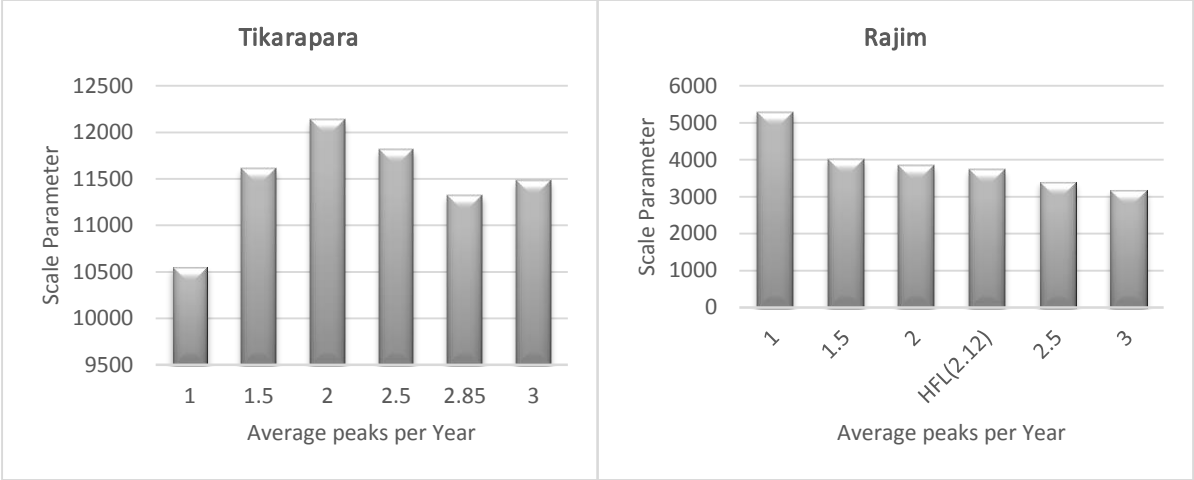
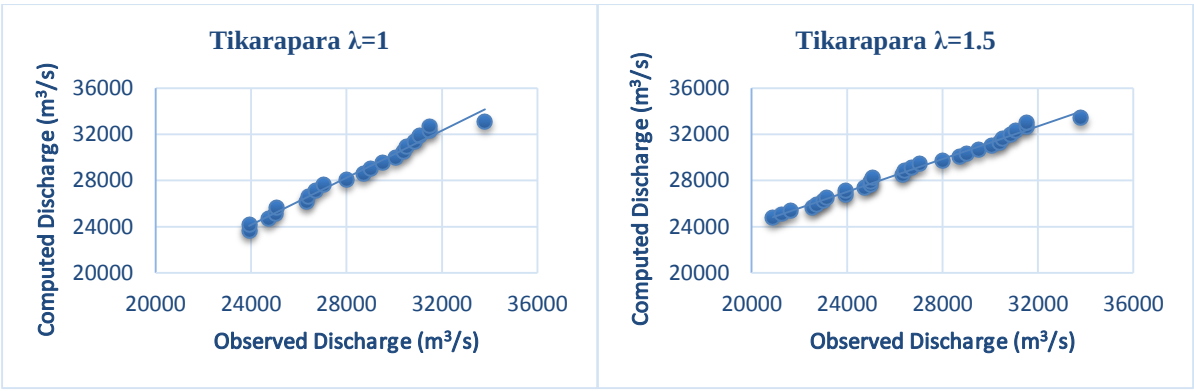
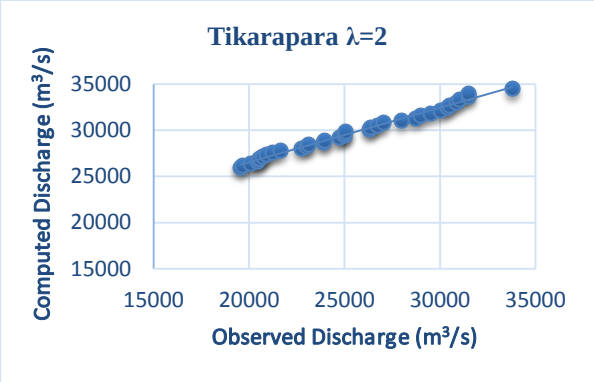
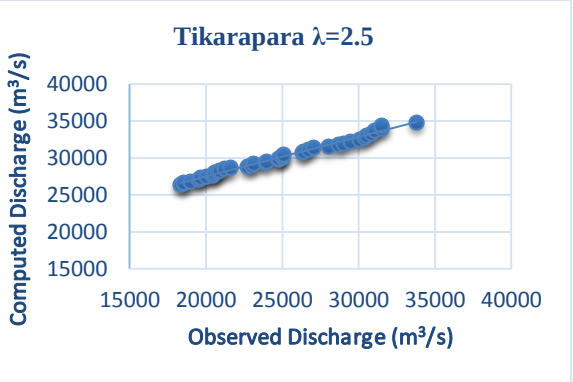


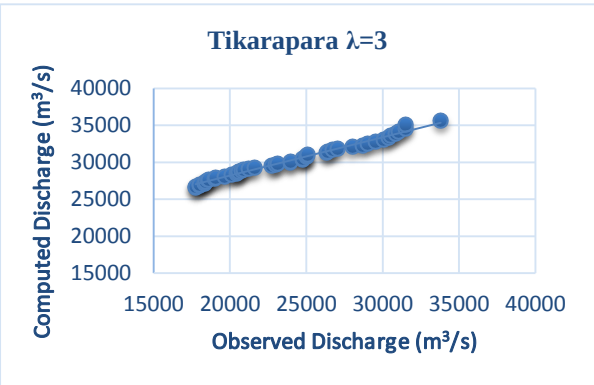
Figure 4.9: Shape and Scale Parameters for the PDS peak floods

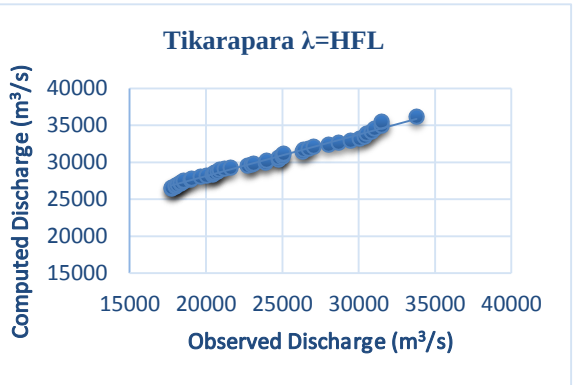
The stability of the GP parameter values has reduced the uncertainty of the quantile estimation by the selection of an appropriate threshold value. Moreover, from a theoretical perspective, the threshold needs to be chosen so that the GP distribution provides a reasonable model for predicting the probability of exceedance for any recurrence interval. A QQ plots for quantile estimates obtained using GP distribution were obtained for all the 22 sampling locations at the different threshold. Figure 4.10 illustrates the plots for Tikarapara and Rajim sampling locations and plots for other stations are presented in Appendix I. Interestingly from the Figure, we can see at lower λ values (i.e., higher threshold), the higher quantiles are under- fitted and at higher λ values (i.e., lower threshold), the models tends to over-fit the higher quantiles. The quantiles obtained from higher return periods are much influenced by the shape (ξ) and scale(σ) parameters of GP distribution model but for smaller return periods it gives good results with change in the threshold values.

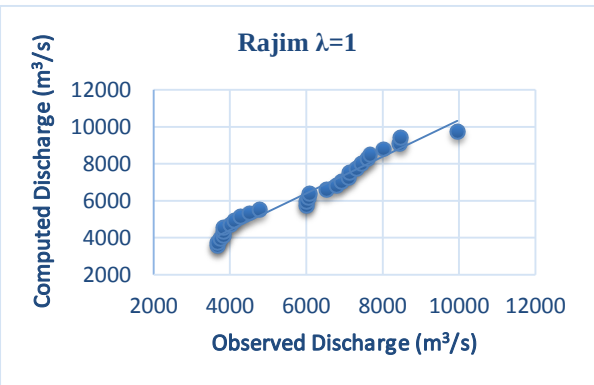


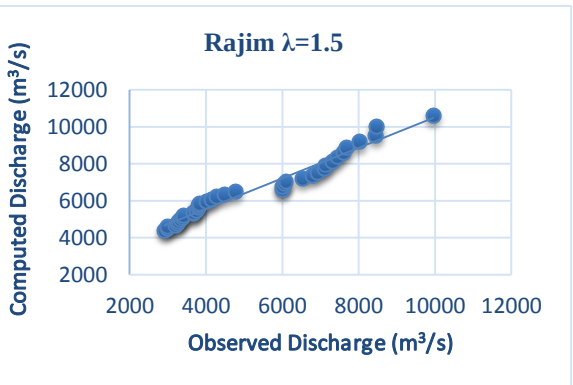

Tikarapara λ=2
Computed Discharge (m³/s)
35000
30000
25000
20000
15000
15000 20000 25000 30000 35000
Observed Discharge (m³/s)

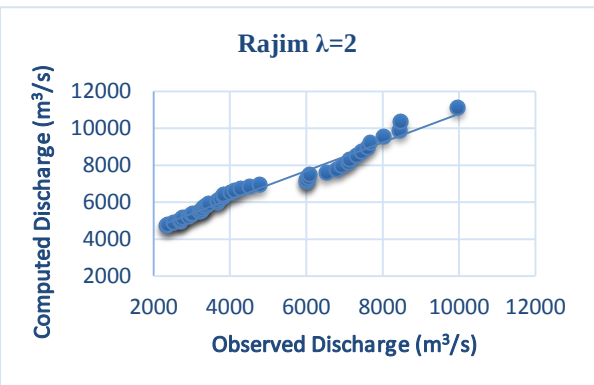

Tikarapara λ=2.5
Computed Discharge (m³/s)
40000
35000
30000
25000
20000
15000
15000 20000 25000 30000 35000 40000
Observed Discharge (m³/s)

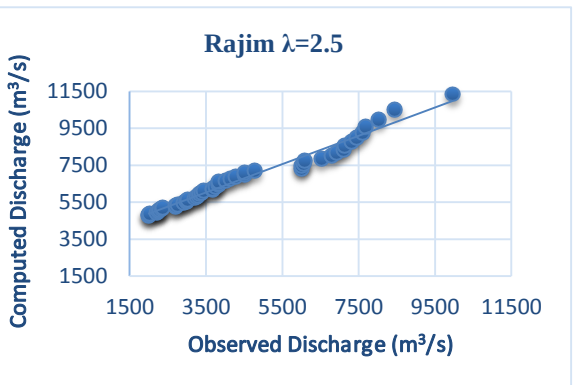

Tikarapara λ=3
Computed Discharge (m³/s)
40000
35000
30000
25000
20000
15000
15000 20000 25000 30000 35000 40000
Observed Discharge (m³/s)


Tikarapara λ=HFL
Computed Discharge (m³/s)
40000
35000
30000
25000
20000
15000
15000 20000 25000 30000 35000 40000
Observed Discharge (m³/s)


Rajim λ=1
Computed Discharge (m³/s)
12000
10000
8000
6000
4000
2000
2000 4000 6000 8000 10000 12000
Observed Discharge (m³/s)


Rajim λ=1.5
Computed Discharge (m³/s)
12000
10000
8000
6000
4000
2000
2000 4000 6000 8000 10000 12000
Observed Discharge (m³/s)


Rajim λ=2
Computed Discharge (m³/s)
12000
10000
8000
6000
4000
2000
2000 4000 6000 8000 10000 12000
Observed Discharge (m³/s)


Rajim λ=2.5
Computed Discharge (m³/s)
11500
9500
7500
5500
3500
1500
1500 3500 5500 7500 9500 11500
Observed Discharge (m³/s)

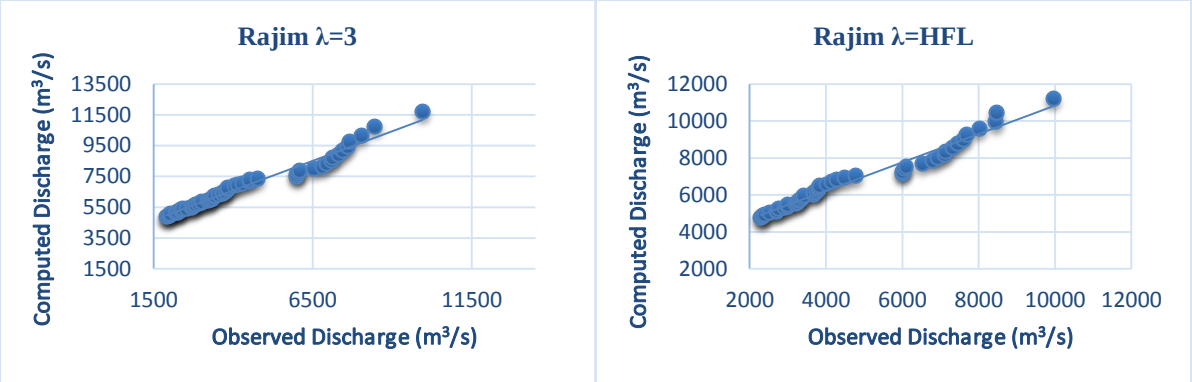


Figure 4.10: QQ Plots at Tikarapara and Rajim Stations for a range of thresholds

Then the performance using relative root mean square error (RRMSE) and relative bias (RBIAS) with the change in the threshold values of GP distribution is evaluated and shown in Figure 4.11. It shows that the performance for different threshold values yields comparable results for shorter return periods, and the difference becomes more noticeable for high return periods with lower threshold values.

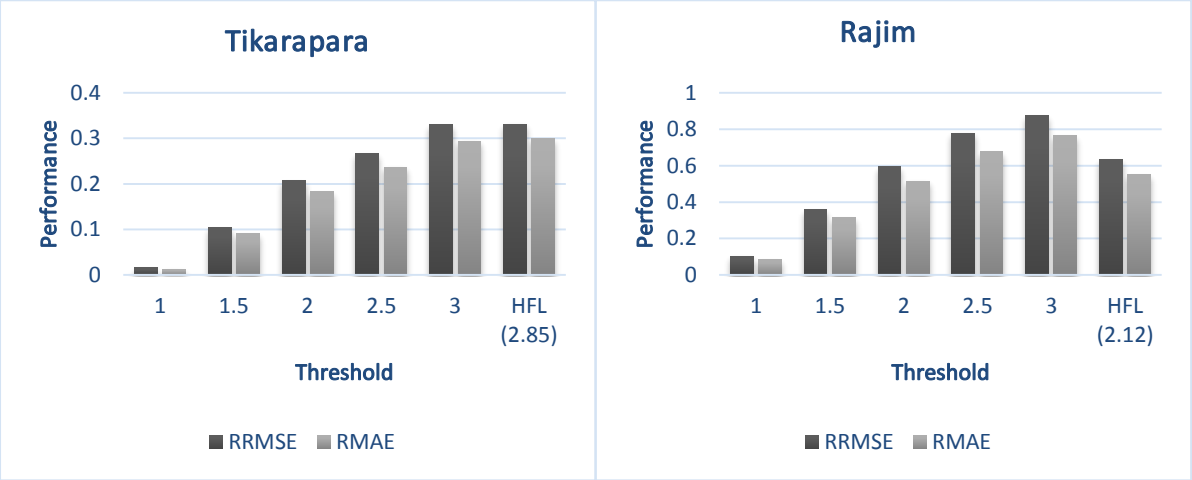


Figure 4.11: Performance evaluation at Tikarapara and Rajim Station for a range of thresholds

4.4 SUMMARY

A framework for selection of the threshold value is provided on the basis of various concepts described in the Chapter and found that the threshold level having average peaks per year (λ) value ranges in between 2 to 3 based on all concepts (i, ii and iii) is suitable to best describe the PDS modeling. Also, from the results, it is concluded that the PDS is random and stationary at most stations. The GP distribution was found be best fitted. The results confirm that the parameter values stabilize with an increase in the threshold. Therefore, in the application of

PDS for the Mahanadi river basin, the threshold value at different stations where the λ values ranging in between 2 to 3 which also include those derived based on the HFL at most stations are preferable. From the performance evaluation almost in all cases less variation in quantile was found with in this range of threshold. Though, the selection of appropriate threshold value for modeling of the GP/PDS was studied at 22 stations across the study area by assuming the whole basin to be statistically homogenous. This assumption signifies a limitation of this study and as proof the assumption we need to develop a regional GP/PDS model based on the selected threshold level however there is only a limited research which is addressed in the next chapter.

CHAPTER-5
REGIONAL FLOOD FREQUENCY ANALYSIS OF PARTIAL DURATION SERIES–CASE STUDY OF MAHANADI BASIN

5.1 INTRODUCTION

On-Site flood frequency analysis is not always possible in the data scare sites to estimate flood of required frequency for hydrologic design. To contend with this situation, hydrologists use regional flood frequency analysis methods that are based on pooling flood information from several watersheds which are similar to the target site watershed in flood producing mechanisms and also, identified the hydrologically homogeneous regions or pooling group. The procedure to identify the homogeneous regions is traditionally referred to as regionalization. Different approaches for identifying the homogeneous region having similar hydrological characteristics called as pooling groups has discussed by Jakob et al. (1999).

Jakob et al., (1999) and Burn, (1990) used region-of-influence pooling methods by using different variables like catchment area, precipitation and soil parameters of the area for grouping of similar gauging sites to make the area becomes homogeneous. Besides, numerous studies have been prepared to findout homogeneous regions based on climatic characteristics (Pearson, 1991), the geographic area [FSR, 1975], and inferring details from gauged to ungauged (Nathan and McMahon, 1990). Investigation of regional flood frequency analysis based on monthly rainfall pattern and geographical proximity was conducted by Gebeyehu (1989) for the Blue Nile River Basin (BNRB). The study had some limitation about the approach that it does not delineate homogeneous regions accurately because the responses of statistical approach in similar rainfall regions are different consequently of changes in basin topography. It was also reported by Cunnane (1989) as well as Rao and Hamed (2000) that stations inside the same geographical homogeneous region cause certain bias in the regionalization.

Trefry et al. (2005) analysed regional flood frequency using PDS series for Michigan State, USA and found that the PDS/GP model gives good results of quantile estimation with λ ranging from 2.2 to 4.07. However, $\lambda^R = 2$ was selected for regional modeling.

In the previous Chapter a framework is provided to select the threshold value based on different concepts and the performance of GP/PDS modeling based on suitability of GP distribution parameters coupled with appropriate threshold value were tested at 22 locations of the study area. Previously, we examined three concepts (i) average number of flood peaks per year (λ=1,

1.5, 2, and so on), (ii) mean exceedance above threshold, and (iii) observed Highest Flood Level (HFL) at each sampling location and corresponding discharge values using Rating Curve, to identify the optimal threshold value which determine the number of upper extremes suitable for PDS analysis and found that the threshold level having average peaks per year (λ) value ranges in between 2 to 3 based on all concepts (i, ii and iii) is suitable to best describe the PDS modeling. However, threshold selection is strongly related to the hypothesis of independence, stationarity, and the combined GP-Poisson assumption. They were tested by, ACF, Poisson assumption with dispersion index, graphical guidance including L-moments ratio diagram, Parameter estimation versus threshold plots, and return period plots.

Here, it is assumed that the whole basin is to be statistically homogenous. Considering above, a regional GP/PDS model based on the selected threshold level has been developed in the Mahanadi river basin. The observed daily discharge data of 22 gauging stations extent over the entire Mahanadi basin was collected from Central Water Commission, Bhubaneswar, and India- WRIS (Water Resources Information System of India). The threshold level having average peaks per year ranges from 2 to 3 at 22 stations were found to be best defined by the GP distribution. In addition, the following fifteen datasets were used for each site: (i) The hydro-meteorological data includes, annual mean precipitation (AMP) of the corresponding site (ii) Physiographical characteristics of the catchments includes area (CA), perimeter (P), mean elevation (E), average slope (SL), drainage density (DD), longest stream length (LSL), compactness ratio (CR), form factor (FF), shape factor (SF), percentage of basin forested (PBF), percentage of basin agricultural (PBA), percentage of basin barren (PBB), percentage of sandy clay loam (PSCL) and percentage of sandy loam (PSL).

5.2 RESEARCH METHODOLOGY AND TECHNIQUE

5.2.1 Regionalization based Flood frequency analysis of the Basin

Generally, in frequency analysis it is crucial that the parent distribution is effective for the whole region and for different return periods (Cunnane, 1988). In the present work an attempt has been made to test the validity of the PDS/GP model in a region using selected threshold value (t) or average peaks per year (λ) used at individual sites in previous Chapter.

The methodology consists of (i) Collection of important input data variables for the study area such as hydrological data, metrological data, topographical and digitized map of the sub-basin to increase understanding of the basin characteristics, or interactions among basin characteristics, (ii) identifying and delineating hydrologically homogeneous region by checking the homogeneity of the entire basin, (iii) identifying the best-fit statistical distributions to the data of each region (iv) esablishing a robust regional frequency curves for

the delineated homogeneous region (v) deriving a relationship between mean POT flood and basin characteristics that can be used for the ungauged catchments within the region and (vi) develop method of estimating quantiles for the ungauged catchments from regional frequency curve for the design of hydraulic structures within the region. However, from the previous chapter it has been confirmed that the PDS data were independent and stationary and can be used for further analysis.

5.2.2 Collection of Important Variables

A selection of variables based on their availability, physical and statistical significance plays an important role. The variables that have been used for regionalization of catchments of study area include: (i) The hydro-meteorological data such as annual mean precipitation (AMP) of the corresponding site (ii) Physiographical characteristics of the catchments include area (CA), perimeter (P), mean elevation (E) , average slope (SL), drainage density (DD), longest stream length (LSL), compactness ratio (CR), form factor (FF), shape factor (SF), percentage of basin forested (PBF), percentage of basin agricultural (PBA), percentage of basin barren (PBB), percentage of sandy clay loam (PSCL) and percentage of sandy loam (PSL).

5.2.3 Discussion on Variables Selection

5.2.3.1 Standardization of Variables

The collected variables have different units of measure hence, the proper conversion is required to confirm that the values are lies between zero and one (Lim and Lye, 2003). Hence, normalization of the variables is needed to make them dimensionless (Jingyi and Hall, 2004):

$$X_{i,0to1} = \frac{X_i - X_{Min}}{X_{Max} - X_{Min}} \tag{5.1}$$

Where X_i = each data point i

X_{Min} = the minimum among all the data points

X_{Max} = the maximum among all the data points

To define the influencing variables, Factor Analysis has been carried out primarily. The main applications of factor analysis techniques are (1) to reduce the number of variables and (2) to detect structure in the relationships between variables that is to classify variables. Therefore, factor analysis is applied as a data reduction or structure detection method. Principal components (PCs) is used as an extraction method in factor analysis to determine the factors from the correlation matrix. We can keep only factors with eigenvalues greater than 1 because any factors having variance less than 1 has less information. This criterion was proposed by

Kaiser (1960), and is one of the most commonly used methods. The variables based on these factors are to be noted as influencing variables by considering the significance of both physical and statistical properties.

The statistical significance of the variables arises from the factor loadings that interpreted the correlations between the respective variables and factors. Factor loadings greater than 0.7 as a cutoff are highlighted which can be the construed correlations between the respective variables and factors.

The physical significance of different variables are also debated. The variable CA expresses the extent of the basin. In large catchment the channel flow phase is more predominant but in small catchments, the overland flow is predominant over the channel flow. Perimeter (P) of the catchments influences the shape of the basin like nearly semicircular shaped catchments contribute high peak whereas elongated catchments contribute low peak. Elevation (E) represents the height above a given level, especially that of the sea. Slope (SL) of a catchment is an important characteristic because it controls the velocity of the flow which provides an indication of the kinetic energy available for water to move toward the basin outlet, and it has been found to be related to total runoff and base flows. Basin having steeper slope gives larger peak discharges. DD is defined as the ratio of the total stream length to the total drainage area and is considered as a significant landscape representative. A large DD makes a condition for rapid removal of runoff that is reflected in a noticeable peak discharge. The LSL characterizes the shape of the catchment for runoff generation which effects time of concentration. The CR is defined as the ratio of the perimeter of the catchment to the perimeter of the circle whose area is equal to that of the basin. The SF is the ratio of the square of catchment length to the catchment area. The FF is the reciprocal of the shape factor. The CR, SF and FF represents the shape of a catchment. Numerically, the values of SF and FF are reciprocal of each other and found significantly different. One of them may be more significant than the other. The land use characteristics such as (PBF, PBA and PBB) rise the penetration and storage capacities of the soils. The catchments with vegetation cover decrease the peak flow, and this effect is noticeable in a small basin. The peak discharge is higher for a dense forest. The soil type includes (PSCL and PSL) effects the infiltration and permeability. Sandy loam has moderate infiltration rates having low runoff potential moderately whereas the clayey loam have low infiltration rates having moderately high runoff potential.

Hall and Minns (1999) used five different variables for regionalization in two regions recognized by the UK Flood Studies Report [National Environment Research Council (NERC) 1975] using the Fuzzy c-means (FCM) algorithm.

5.2.3.2 Selection of Suitable Number of Variables

Selection of suitable independent variables plays an important role in the regional flood frequency analysis hence, an effective, small number and influential variables is considered for further analysis based on the importance of physical characteristics and the best correlation with the dependent variable. In this concern, multiple regressions analysis plays an important role for selecting the particular number of variables.

5.2.4 Identification and Delineation of Hydrologically Homogeneous Region

5.2.4.1 Cluster Analysis

Cluster analysis (CA) divides sites into groups based on the statistical distance which reflect the similarity (or dissimilarity) among a set of attributes. Various methods of cluster analysis are available in the literature on statistics, including hierarchical approaches such as single linkage, complete linkage, average linkage and Ward's method, as well as non-hierarchical approaches such as the k-means method. Above methods have been extensively applied for delineation of homogeneous regions (Burn, 1989, 1997, 2000; Baeriswyl and Rebetez, 1997; Hosking and Wallis, 1997; Chiang et al., 2002a; Castellarin et al., 2001; Rao and Srinivas, 2006). In this study, two methods of classification were applied for identification of homogeneous regions, such as Hierarchical Cluster and K- Means Cluster using ARC-GIS toolbox.

Hierarchical Cluster (HC)

Hierarchical Cluster considers a graphical representation of a matrix of distances represented as a structure like a tree called a dendrogram or tree – where the objects are joined together in a hierarchical fashion from the closest, that is the most similar, to the farthest apart, that is the most different. The purpose of this algorithm is to join together objects into successively larger clusters, using some measure of similarity or distance. The most straightforward way of computing distances between objects in a multi-dimensional space is to compute Euclidean distances. This is probably the most commonly chosen type of distance. It simply is the geometric distance in the multidimensional space. It is computed as:

$$dis\tan ce(x,y)=\left\{\sum_{i}(x_i-y_i)^2\right\}^{1/2} \quad (5.2)$$

When considering the Euclidean distance, the scale of the dimension of the variables is an issue, because when the dimension of the scale is changed then the distance between objects is affected.

K-Means Clustering Methods (KM)

K-Means is one of the simplest unsupervised algorithms which classify a given feature into k clusters by defining k centers for each cluster to minimize the differences among the features. It uses NP (nondeterministic polynomial time)-hard algorithm, a greedy heuristic is working to group structures. In these methods, the desired number of clusters is specified in advance, and the 'best' solution is chosen. In K-Means, initially, recognization of seed features for each group is carried out, and the first seed is selected randomly. Also, the rest of the seeds follow the random process by applying a weighting factors helping for selection of seed which is present far away from the existing seed features, and this process is known as K Means ++. Once the seed features are recognized, all features are given to the closest seed feature and, a mean data center is calculated, and each feature is reassigned to the nearest center. The process of calculating an average data center for each cluster and then reassigning features to the nearest center endures until the group becomes stable (up to a maximum of 100 iterations).

According to Thandaveswara and Sajikumar (2000), clustering process has some objectives namely, (i) a cluster has sufficient data for hydrological studies and (ii) the data is statistically homogeneous in nature.

In this study, the number of clusters which make the region homogeneous is decided based on (i) the Hierarchical Cluster supported by K- Means Cluster, (ii) Based on Robson and Reed (1999), which says that a group should have at least seven sites, and (iii) the number of stations in a group follows the 5T rule i.e, the five times the return period for attaining a reasonable return period.

Further, a statistical homogeneity test is used to verify the performance of different clustering techniques to measure the amount of inconsistency inside the regions. Additionally, to prove the method, a comparison between the scale and the dispersion values of the L-moment (LCv) and conventional moment (Cv) of gauging stations of different regions is performed.

5.2.4.2 Statistical Homogeneity Tests

The tests used in this study are (a) discordance measure tests, (b) Cv and LCv-based homogeneity test and, (c) the comparison based on statistical analysis for the previously selected threshold level.

5.2.4.2.1 Discordance Measure

The discordance measure is a measure that identifies sites that are grossly discordant with other locations in the same region. It is a useful measure in assessing whether any of the regions obtained from the cluster analysis contain potential outliers and should, therefore, be adjusted accordingly. The discordance measure D evaluates the distance of each site on the basis of statistical properties.

If $U_i = \left[t^{(i)}, t_3^{(i)}, t_4^{(i)}\right]^T$ is the vector containing the t, t_3 and t_4 values for site (i), then the group average for NS (number of sites) within the region is given by

$$\overline{U} = \frac{1}{NS}\sum_{i=1}^{NS} U_i \tag{5.3}$$

The sample covariance matrix is given by

$$S = (NS-1)^{-1}\sum_{i=1}^{NS}\left(U_i - \overline{U}\right)\left(U_i - \overline{U}\right)^T \tag{5.4}$$

The discordance measure (D_i)is defined as:

$$D_i = \frac{1}{3}\left(U_i - \overline{U}\right)^T S^{-1}\left(U_i - \overline{U}\right) \tag{5.5}$$

A station is said to be a discordant, when the value of $D_i \geq 3$.

5.2.4.2.2 Cv and LCV- based Homogeneity Test

In this test, an assumption is made based on the statistical similarity. To examine the assumption (Lettenmaier and Potter, 1985; Cunnane, 1989), the mean values of coefficient of variation (Cv) and the on-site coefficient of variation (CC) are calculated using both conventional and L-moments methods of the region and found that higher values of both Cv and CC results in lowering the performance. Hence, the value of CC should be retained low, generally < 0.30 for the well performance of the index flood method. Here, we calculated Cv, LCv, and their respective CC values using L-moments method. The procedure is described below.

i) For each site, in a region calculate mean, standard deviation and coefficient of variation Cv

$$\overline{Q}_i = \frac{\sum_{j=1}^{n_i} Q_{ij}}{n_i} \quad (5.6)$$

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{n_i} \left(Q_{ij} - \overline{Q_i}\right)^2}{n_i - 1}} \quad (5.7)$$

$$CV_i = \frac{\sigma_i}{\overline{Q}_i} \quad (5.8)$$

Where Q_{ij} = flows rate for station j in region i

Cv_i= Coefficient of Variation for site i

For calculation of LCV use,

$$LCv_i = \frac{L_{2i}}{L_{1i}} \quad (5.9)$$

Where LCv_i is the dimensionless coefficient of variation calculated from L- moments

ii) For each region, using the statistics calculated in step 1, compute the regional mean, Cv and LCv; standard deviation of Cv and LCv, and finally the corresponding CC using the following relations

$$\overline{CV} = \sum_{i=1}^{N} \frac{CV_i}{N} \quad (5.10)$$

$$\sigma_{CV} = \sqrt{\frac{\sum_{i=1}^{N} \left(CV_i - \overline{CV}\right)^2}{N}} \quad (5.11)$$

$$CC = \frac{\sigma_{CV}}{CV} \quad (5.12)$$

Where N is the number of sites in a region

The region is declared to be homogenous if CC<0.30.

5.2.4.2.3 Statistical Comparison

Statistical comparison of different stations that belong to the same region are analyzed by plotting the LCv values of peak over threshold flow against LCs and LCk values.

5.2.5 Selection of Regional Distribution

This part of the study deals with selecting and validating of the most suitable frequency distribution from on-site data as well as making it valid for a statistically homogenous region.

For a selection of suitable regional frequency, Goodness-of-Fit tests including Z-statistic and the L-moment ratio diagram (LMRD) are calculated.

The regional weighted average L- moment ratios according to record length used to classify a regional frequency distribution apart from on-site data can be calculated as follows:

If there are N sites in a region with sample size n_1, n_2,.. n_N, respectively and the sample L-moment ratios at site i are denoted as t^i_r with the r-order of L- moment ratios then the regional weighted average L- moment ratios are defined:

$$t_r^R = \frac{\sum_{i=1}^{N} n_i t_r^i}{\sum_{i=1}^{N} n_i} \tag{5.13}$$

A regional distribution will be examined by Z-statistic and the L-moment ratio diagram.
The Z-statistic statistics specify the appropriateness of a candidate distribution to a series of datasets. It also shows how to fit the L-moment ratios of the candidate distribution match with the regionally averaged L-moment ratios. Theoretically, regional distribution is best when Z value is the smallest being close to zero. The Z-statistics is represented by

$$Z^{DIST} = \left(Z_4{}^{DIST} - Z_4 + B_4\right)/\sigma_4 \tag{5.14}$$

Where DIST = a particular distribution; $Z_4{}^{DIST}$ = L-kurtosis for fitted distribution; Z_4 = pooled L-kurtosis; B_4 = bias correction; and σ_4 = estimate of sample variability of L-kurtosis. The Z^{DIST} value should be close to zero or has a minimum value. However, a value between -1.64 and +1.64 are suitable for a fitting distribution at 10% significance level.

LMRD with five candidate distribution models namely, Generalised Pareto (GP), Generalised Extreme Value (GEV), Log Normal (LN), Pearson Type III (P3), and Generalised Logistic (GL) which are arguably alternatives for fitting PDS series is plotted to find the suitable distribution for all regions (Jaiswal et al., 2003). For a given region, a suitable distribution is selected by plotting the sample regional LCv, LCs, and LCk, respectively, on the LMRD.

5.2.6 Derivation of Regional Frequency Curve for the Homogeneous Regions

The regional flow frequency curve is developed by using the following relationships:
(i) Relationship between index flood and return period T, and

(ii) Relationship between index flood and catchment characteristics.

(i) The regional flood frequency curve is established for each region based on suitable distribution to calculate the deviations in the standardized flow of various return periods. Actually, uncertainty reduction in predicted quantiles at ungauged sites is of major importance in regional flood frequency analysis. The index flood method has a hypothesis that floods from different catchments within a region normalized by their mean annual flood from a single distribution. Hence, normalization of the flood data is required for further analysis and express as:

$$X_i = \frac{Q_i}{\overline{Q}} \tag{5.15}$$

Where $\overline{Q} = \frac{1}{N}\sum_{i=1}^{N} Q_i$

Q_i = peak over threshold for year i

X_i= standardized peak over threshold for year i

The flow quantile Q_T, is estimated as

$$Q_T = \overline{Q}\tilde{X}_T \tag{5.16}$$

Where Q_T = POT flow for T year, return period

$\overline{Q}$ =mean of POT flow and $\tilde{X}_T$ = standardized quantile or growth factor for T year return period.

$$X_T = \mu + \frac{\sigma}{\xi}\left[1 - \left(\frac{1}{\lambda T}\right)^{\xi}\right] \tag{5.17}$$

where λ is the number of average peaks in per year, location (μ), scale(σ), and shape (ξ) parameters .

(ii) The index flood estimation method is one of the best methods for average peak flood estimation in the homogenous regions. The method is divided into two categories i.e., direct and indirect methods. The direct method estimates the index flood from observed data at the interested sites and the indirect method estimates the index flood from other hydrological, meteorological, and geomorphological features. Bocchiola et al. (2003) figured out various methods of index flood estimation. Brath et al. (2001) presented that the regression analysis is used for calculating index flood at ungauged sites in a hydrological homogeneous region, which include the multiple regression relationship in between index flood and catchment characteristics. Initially, a correlation analysis was done between the catchment features and

the index flood for each region of the study area. Then regression analysis was carried out by using two or more independent variables, and the regression model was selected on the basis of calculated correlated values.

5.3 RESULTS AND DISCUSSION

5.3.1 Summary of partial duration series (PDS) and its frequency distribution

For this study, we used three extracted PDS series denoted as PDS2, PDS2.5 and PDS3 for further analysis because from the previous study we concluded that the selected threshold level (t) having λ ranges in between 2 to 3 is the best to describe the PDS modeling at all 22 stations. These extracted peaks were confirmed to be random and stationary from the analysis of the previous study. Also, the result reveals that GP distribution best describes these peaks at most of the stations using L- moment ratio diagram approach combined with the GP distribution parameters and quantile estimation.

5.3.2 Collection, Standardization and Selection of Important Variables

The detailed collection of important variables is discussed in Section 5.2.2. The normalization of the important variables using equation (5.1) is essential before applying any methods for analysis. In the present analysis we used fifteen variables such as: CA, Perimeter, Elevation, Avg Slope, DD, CR, PBF, PBA, PBB, PSCL, PSL, AMP, LSL, FF and SF for each catchments.

Principal components (PCs) method is applied to the above fifteen variables, which is as an extraction method in factor analysis to calculate the variances concerning PCs and are as presented in Table 5.1. Table shows the five PCs having eigenvalues greater than 1, are selected. It explains nearly 85% of the variance. The statistical significance of the variables are calculated from the factor loadings which interpret the correlations between the respective variables and factors. The loadings of individual factors are revealed in Figures 5.1(a) to 5.1(e), and their values are noted in Table 5.2. Factor loadings greater than 0.7 and less than -0.7 as a cutoff, are highlighted as they represent the most important information for the interpretation of factors.

From Table 5.2, it is shown that (a) CA, P, CR and LSL dominate FC1; (b) PBA, E, SL,PBF dominate FC2; (c) DD, PBB dominate FC3, while FC5 is dominated by AMP and; (d) no loadings are found to dominate FC4. Here FC1, FC2, FC3, FC4 and FC5 are the 1^{st} to 5^{th} Factored Principal Components.

Table 5.1: Result of Factor Analysis

Principal Components	Eigen value	% Total variance	Cumulative Eigenvalue	Cumulative %
1	4.680130	29.25081	4.68013	29.25081
2	3.775154	23.59471	8.45528	52.84552
3	2.245262	14.03289	10.70055	66.87841
4	1.923509	12.02193	12.62405	78.90034
5	1.035428	6.47142	13.65948	85.37177

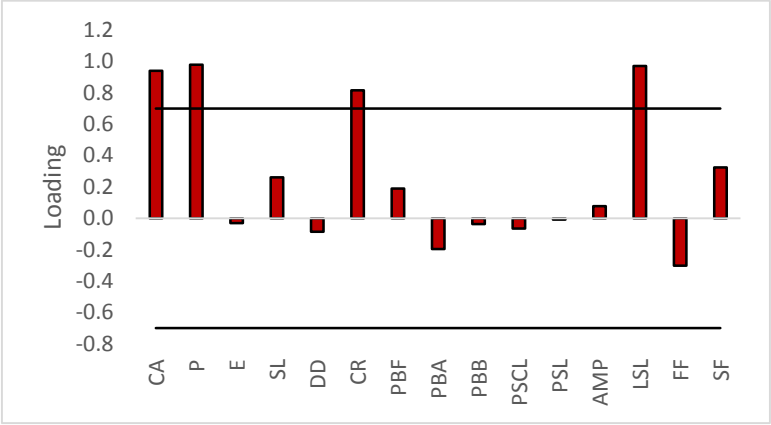


Figure 5.1(a): Loading of different variables of FC-1

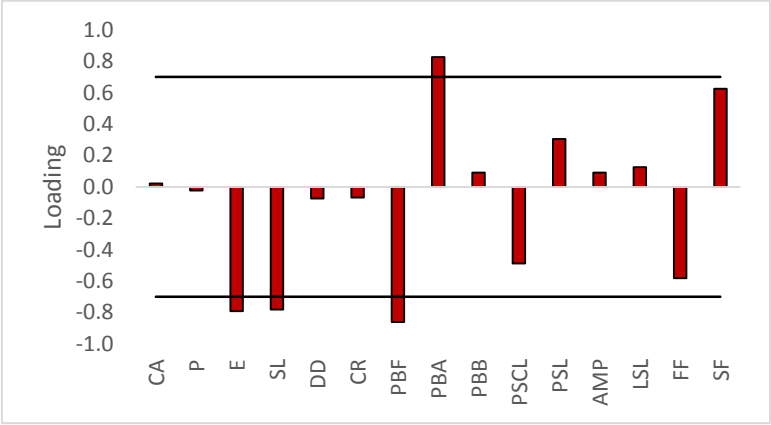


Figure 5.1(b): Loading of different variables of FC-2

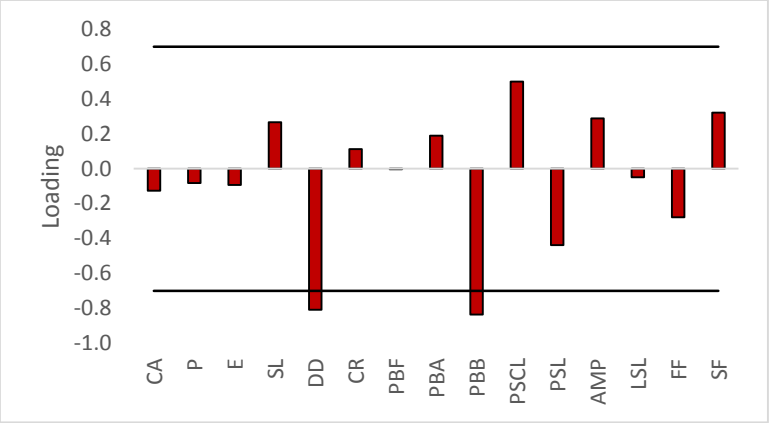


Figure 5.1(c): Loading of different variables of FC-3

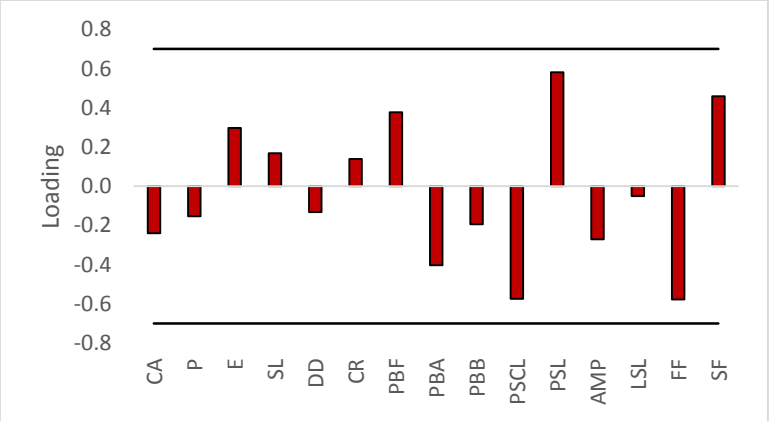


Figure 5.1(d): Loading of different variables of FC-4

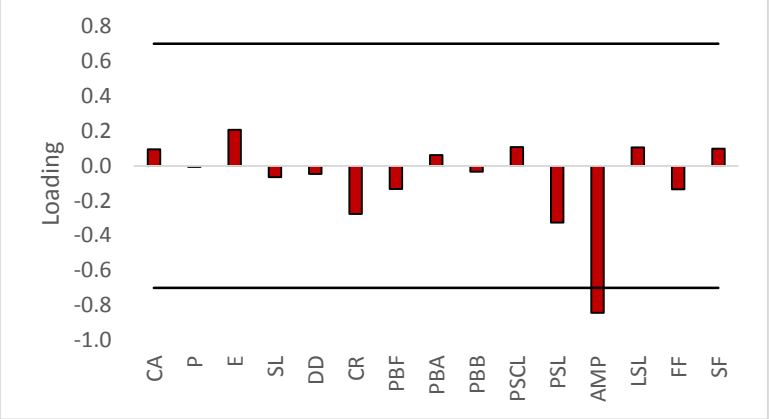


Figure 5.1(e): Loading of different variables of FC-5

Table 5.2: Loading values of Five FCs

VARIABLES	FC1	FC2	FC3	FC4	FC5
CA	0.939770	0.021330	-0.126948	-0.239768	0.094117
P	0.978869	-0.022417	-0.081649	-0.153051	-0.006153
E	-0.031506	-0.792562	-0.092262	0.297196	0.206599
SL	0.261299	-0.782997	0.267103	0.168688	-0.065277
DD	-0.086055	-0.073613	-0.809949	-0.131386	-0.047617
CR	0.816277	-0.068188	0.111894	0.139222	-0.276903
PBF	0.190225	-0.861462	-0.002972	0.376946	-0.132520
PBA	-0.195752	0.826982	0.189754	-0.403089	0.060882
PBB	-0.036083	0.091725	-0.836483	-0.193394	-0.034448
PSCL	-0.066487	-0.488051	0.500124	-0.574351	0.107110
PSL	-0.009018	0.306112	-0.437597	0.580860	-0.325957
AMP	0.076355	0.091588	0.288962	-0.270954	-0.843572
LSL	0.971239	0.126151	-0.049540	-0.049586	0.105999
FF	-0.302315	-0.582862	-0.277609	-0.577405	-0.134770
SF	0.324123	0.626097	0.322847	0.458450	0.098809

Hence, from the loadings of the factor analysis, eleven variables such as CA, P, CR, LSL, E, SL, PBF, PBA, DD, PBB and AMP are found to be suitable for further analysis.

When the number of independent variables are large, small number of effective variables based on the physical process and good correlation with the dependent variable are selected for further analysis. Hence, to reduce the number of variables, multiple regression analysis between the independent variables with the dependent variable (average Q_{pot}) is carried out. On regressing with fifteen variables, the R^2 value for extracted PDS varies in between 0.971 to 0.973. By reducing one inferior variable regularly from selected list, the R^2 is reduced. With eleven variables (viz., CA, P, CR, LSL, E, SL, PBF, PBA, DD, PBB and AMP), the R^2 value varies from 0.907 to 0.911, with ten variables (viz.,CA, CR, LSL, E, SL, PBF, PBA, DD, PBB and AMP), the R^2 value varies in between 0.899 to 0.901 and with nine variables (CA, CR, LSL, SL, PBF, PBA, DD, PBB, and AMP), the R^2 value varies in between 0.879 to 0.882. Even with nine variables, the information loss is very less. So, for further analysis, nine variables (viz.,CA, CR, LSL, SL, PBF, PBA, DD, PBB and AMP) are also used in addition to 11 and 15 variables.

5.3.3 Identification and Delineation of Hydrologically Homogeneous Region

After selection of a suitable number of variables, the analysis is started for the number of clusters. The hierarchical clustering based on Ward's method is applied for nine, eleven and

fifteen variables. Hierarchical Clustering (HC) method with all variables is shown in Figures 5.2 (a)-5.2(c).

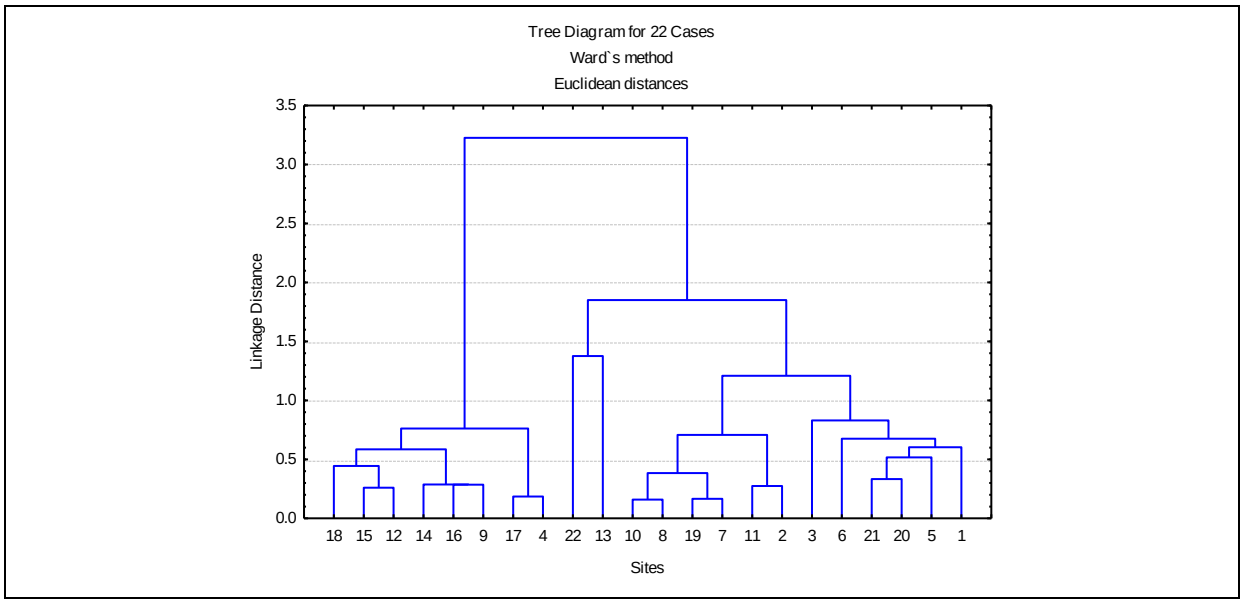


Figure 5.2 (a): Result of Hierarchical Clustering using nine variables (dendrogram)

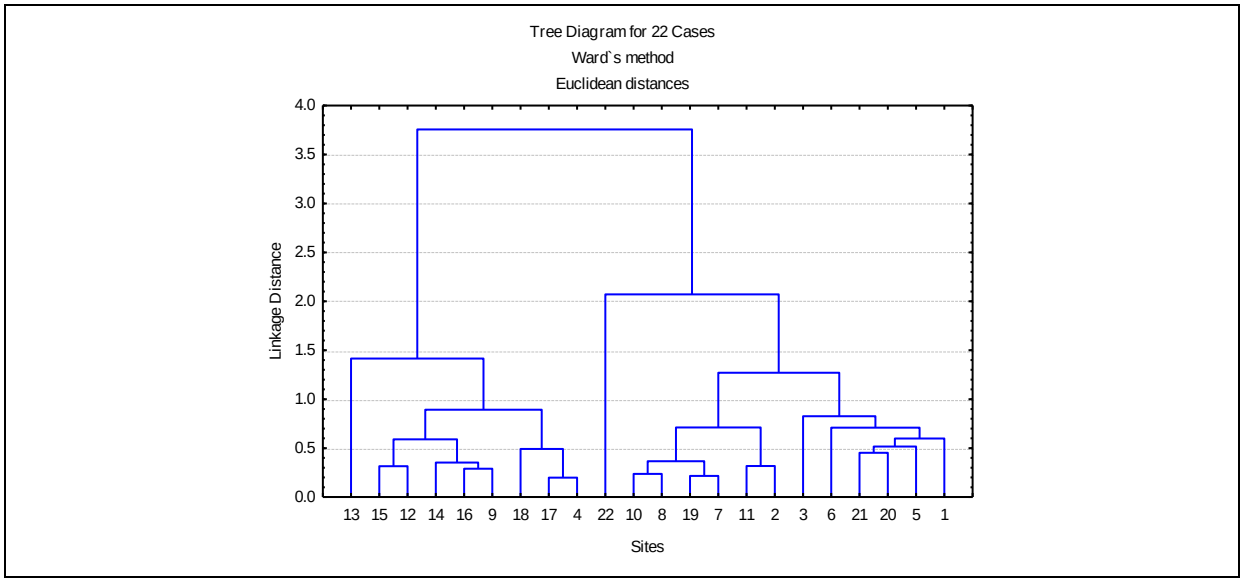


Figure 5.2 (b): Result of Hierarchical Clustering using eleven variables (dendrogram)

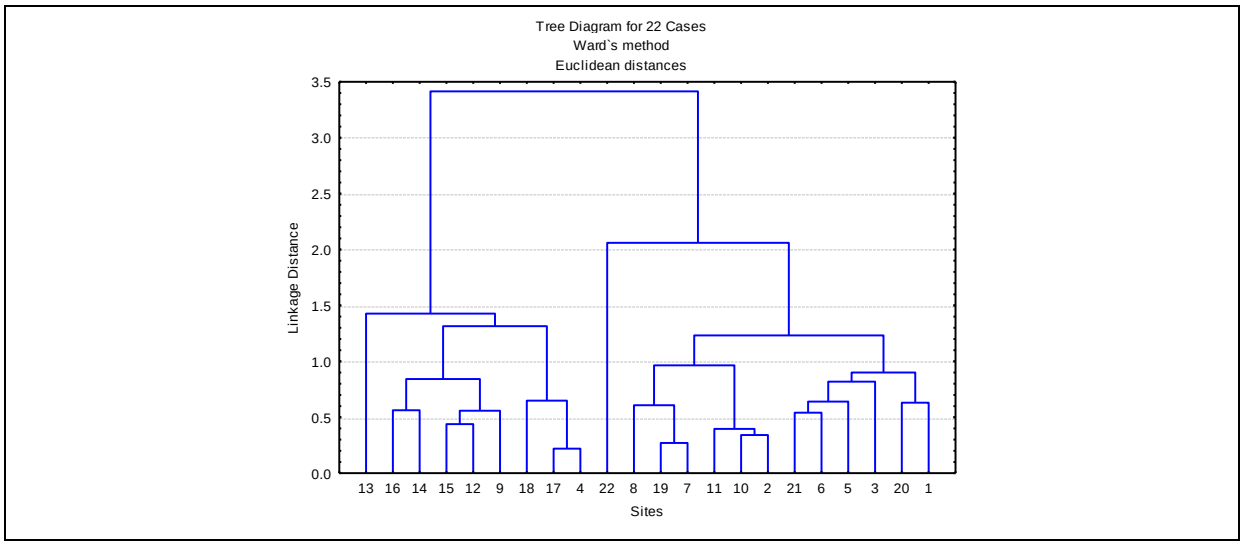


Figure 5.2 (c): Result of Hierarchical Clustering by using fifteen variables (dendrogram)

A larger number of clusters would have resulted in disadvantages, hence, a small number of the cluster with fewer sites and fewer station years are considered for finding homogeneous region. For verification, two clusters are tried with nine, eleven and fifteen variables by using K-means clustering method using grouping analysis tool. The delineation of homogenous regions is performed by using the geographical information system (GIS) software ArcView. All sample stations are located on a digitized map by latitude and longitude. For each station, the statistical values (LCs and LCk) are computed. It is assumed that the LCs and LCk values of one station vary linearly with the neighboring stations. The distance between stations is determined from geographical coordinates and the LCs and LCk values are linearly interpolated to locate the boundaries between two stations of different regions. A small script is used to demarcate the region boundary exactly at the midpoint of the two borders, obtained from linear interpolation of LCs and LCk. It is the computed means across clusters (which is useful for visually summarizing the differences in means between clusters) and F-test (which can compare the within-cluster variability (small if the classification is good) to the between-cluster variability (large if the classification is good)) for each variable) as shown in Table 5.3.

Table 5.3: Analysis of Variance with different Variables

SL NO.	VARIABLES	FIFTEEN VARIABLES		ELEVEN VARIABLES		NINE VARIABLES	
		F test	**p-levels**	**F test**	**p-levels**	**F test**	**p-levels**
1	CA	1.238	0.2789	1.23870	0.278921	1.23870	0.278921
2	SL	37.955	0.000005	37.95576	0.000005	37.95576	0.000005
3	DD	0.00194	0.965264	0.00194	0.965264	0.00194	0.965264
4	CR	2.36925	0.139419	2.36925	0.139419	2.36925	0.139419

5	PBF	33.64671	0.000011	33.64671	0.000011	33.64671	0.000011
6	AMP	0.02992	0.864412	0.02992	0.864412	0.02992	0.864412
7	PBA	25.11651	0.000067	25.11651	0.000067	25.11651	0.000067
8	LSL	0.95499	0.340127	0.95499	0.340127	0.95499	0.340127
9	PBB	0.98403	0.333056	0.98403	0.333056	0.98403	0.333056
10	E	26.8860	0.000045	26.88607	0.000045		
11	P	1.80064	0.194673	1.80064	0.194673		
12	PSCL	0.20760	0.653562				
13	PSL	0.00360	0.952731				
14	SF	0.82122	0.375614				
15	FF	0.16764	0.686566				

The allocation of different sites and means for both clusters of nine variables are shown in Figures 5.3(a) - 5.3(b). Box plot for both clusters have been shown in Figures 5.4(a) - 5.4(b).

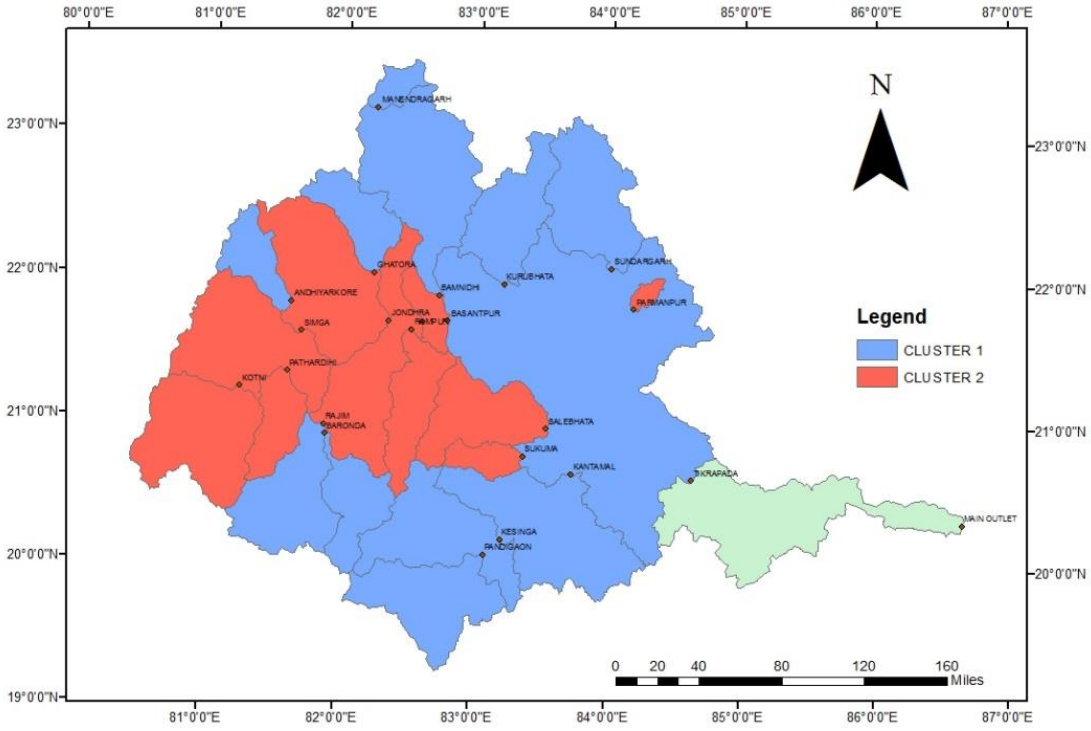


Figure 5.3(a): Result of K- means Clustering using nine variables

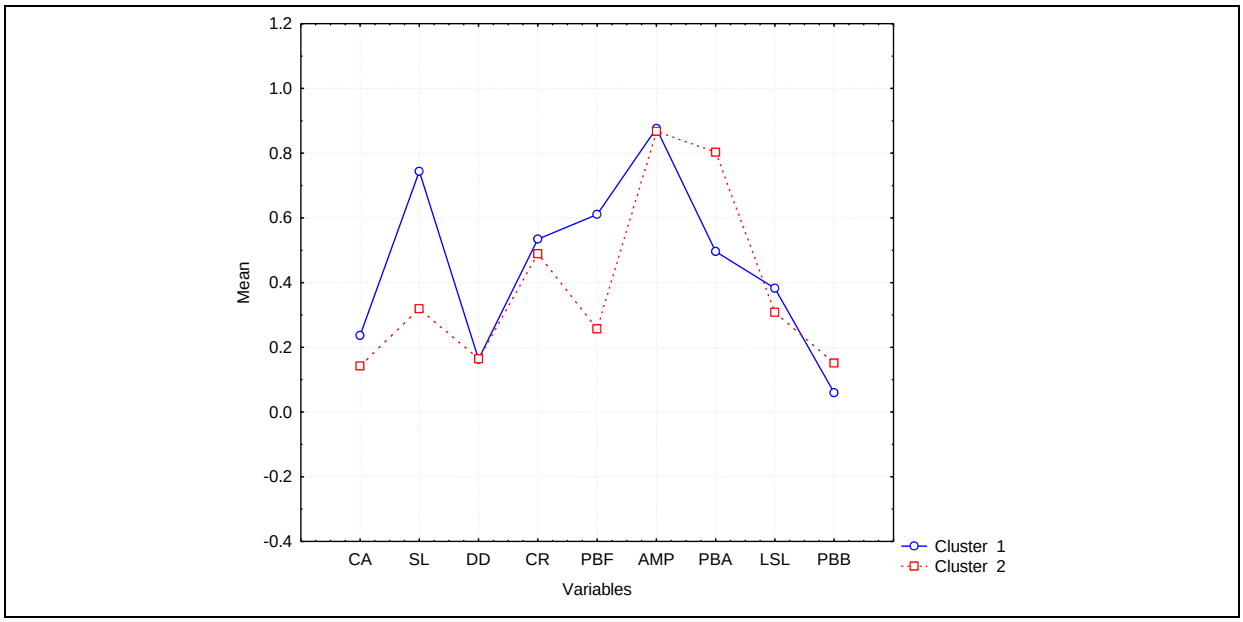


Figure 5.3(b): Mean plots of cluster by using nine variables

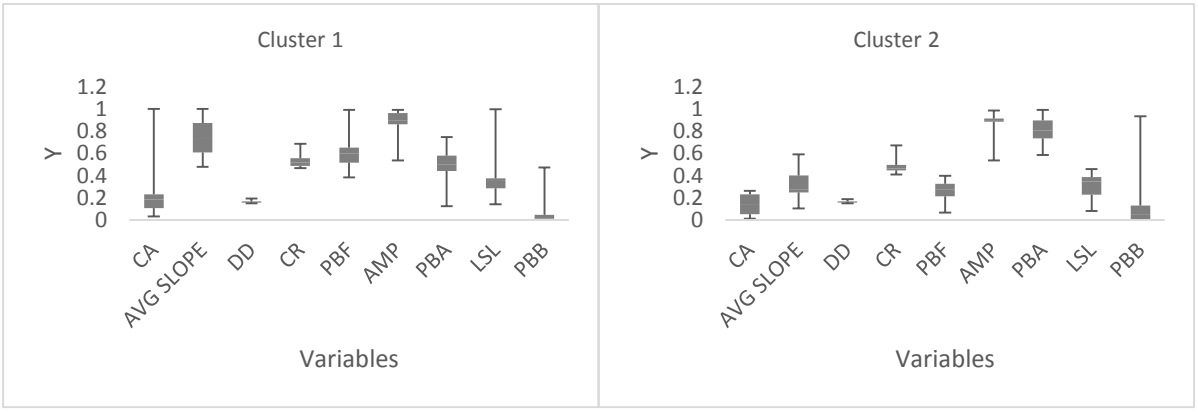


Fig. 5.4: Box plot of variables: (a) Cluster-1; (b) Cluster-2

It has been observed that the result obtained from HC method does not remain reliable with the variation in variables whereas K-means (KM) method results remain consistent for different set of variables. Judging from Table 5.3 the magnitude (and significance levels) of the F values, variables SL, PBF, PBA and E are the major criteria for assigning sites to clusters. From the Figure 5.3(b) and Figures 5.4 (a)- 5.4(b), it is found that the sites in Cluster 2 have (i) small CA, (ii) area having smaller slope, (iii) area having less percentage of forested basin, (iv) area having high percentage of agriculture, (v) area having highest barren and, (vi) area about the same DD, CR, and AMP. Cluster 1 belong to coastal basin.

Table 5.4 demonstrate the sampling sites assigned to each cluster by using 9, 11 and 15 variables. The clustering has been complete effectively with a minimum of seven sites in a

group (Robson and Reed, 1999). Another one is the number of stations in a group follows the 5T rule i.e., five times the return period for attaining a reasonable return period. Here, Cluster-1 has a station year of 370, and Cluster-2 has 283. Therefore, based on 5T rule, both the clusters give better result of up to 56 and 75 years return period using the current distribution. The station Simga is appeared in cluster 1 by using the HC method with nine variables and is appeared in cluster 2 by using both the methods with eleven and fifteen variables because HC method use the Ward's minimum variance criterion that minimizes the total within-cluster variance. Once confirmed, the KM method with nine variables result is accepted for final clusters.

Table 5.4: Sites allocated to each cluster by using 9, 11 and 15 variables

Clustering Method	**No. of Variables Considered**					
	Nine		Eleven		Fifteen	
	Cluster-1	Cluster-2	Cluster-1	Cluster-2	Cluster-1	Cluster-2
HC	1,2,3,5,6,7,8,10,11,13,19,20,21,22	4,9,12,14,15,16,17,18	1,2,3,5,6,7,8,10,11,19,20,21,22	4,9,12,13,14,15,16,17,18	1,2,3,5,6,7,8,10,11,19,20,21,22	4,9,12,13,14,15,16,17,18
KM	1,2,3,5,7,8,10,11,19,20,21,22	4,6,9,12,13,14,15,16,17,18	1,2,3,5,7,8,10,11,19,20 ,21,22	4,6,9,12,13,14,15,16,17,18	1,2,3,5,7,8,10,11,19,20,21,22	4,6,9,12,13,14,15,16,17,18

5.3.4 Homogeneity Tests

To verify the performance of the above clustering techniques, various statistical homogeneity tests are used such as: (a) discordance measure tests using equations (5.3-5.5), (b) Cv and LCv-based homogeneity test using equations (5.6-5.12) and, (c) statistical comparison. According to this test, all stations except Manendragarh of Cluster 1 and Paramanpur of Cluster 2 satisfy homogeneity criteria. Results computed in a group of different sites are summarized in Table 5.5 and 5.6. The value of Di varies from station to station. In most of the sites, the extracted PDS3, Di values are generally lower than the PDS values for other λ values. There are two stations viz; Manendragarh and Paramanpur with their Di values approaches to three. The conventional Cv for Cluster 1 is greater than 0.3 and for Cluster 2 approaches to 0.3 signifying a slightly less suitable stations in contrast with other stations. These sites are discordant may be due to the existence of inaccuracies in data or some other local conditions. So, these two stations are not used for further analysis.

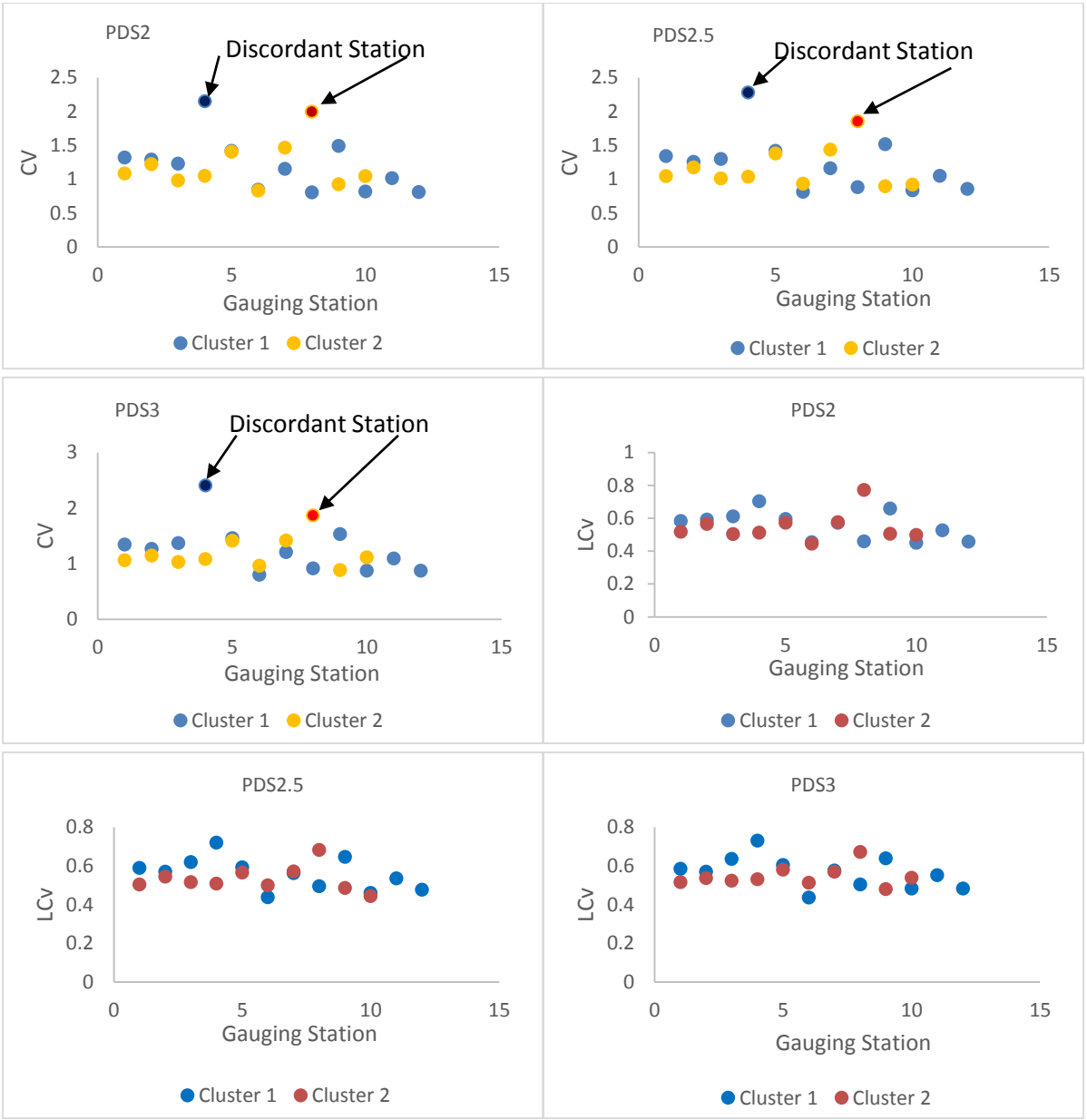
Table 5.5: Discordance ratio (Di) computed at stations for the Regions in Mahanadi River

SL NO	SITE	PDS2 (DI)	PDS2.5 (DI)	PDS3 (DI)	CONCLUSION
1	Andhiyarkore	0.2518	0.3184	0.1891	Homogeneous
2	Bamnidhi	0.4101	0.2253	0.1058	Homogeneous
3	Baronda	1.0764	1.2211	1.1639	Homogeneous
4	Basantpur	0.3321	0.2294	0.3232	Homogeneous
5	Ghatora	0.3432	0.3335	0.3220	Homogeneous
6	Jondhra	0.6612	0.3003	0.3232	Homogeneous
7	Kantamal	0.2230	0.4514	1.4655	Homogeneous
8	Kesinga	0.4262	0.5281	0.2849	Homogeneous
9	Kotni	0.5783	0.5482	0.4797	Homogeneous
10	Kurubhata	2.2286	2.1967	1.9096	Homogeneous
11	Sundargarh	1.1052	0.8307	0.5703	Homogeneous
12	Sukuma	0.4323	0.5326	1.2824	Homogeneous
13	Simga	0.0595	0.1602	0.3271	Homogeneous
14	Seorinarayan	1.8048	2.7670	1.1908	Homogeneous
15	Salebhata	0.6253	1.1288	1.7880	Homogeneous
16	Rampur	0.2972	0.4479	0.4516	Homogeneous
17	Pathardihi	2.7274	1.3465	0.4574	Homogeneous
18	Paramanpur	2.7274	2.5385	2.5874	Heterogeneous
19	Pandigaon	1.3101	1.8498	2.0827	Homogeneous
20	Manendragarh	2.9113	2.8173	2.6531	Heterogeneous
21	Rajim	0.5984	0.1307	0.2814	Homogeneous
22	Tikarapara	1.1152	1.0963	0.9711	Homogeneous

Table 5.6: Cv and LCv–Based Homogeneity Tests for the Clusters in Mahanadi River

Cluster	PDS2 CC	PDS2.5 CC	PDS3 CC	Conclusion	PDS2 CC	PDS2.5 CC	PDS3 CC	Critical CC Value	Conclusion
	Conventional Cv-based				LCv based				
1	0.3264	0.3357	0.3441	Heterogeneous	0.1538	0.1479	0.1446	0.30	Homogeneous
2	0.2866	0.2601	0.2416	Homogeneous	0.1616	0.1220	0.0963	0.30	Homogeneous

The Cv and LCv values for all three PDS i.e PDS2, PDS2.5 and PDS3 are plotted for both clusters against LCs and LCk values to know the statistical nature of the grouped stations that form the homogeneous region is shown in Figure 5.5. It is observed that stations of similar region confirm separate layers, like Cluster 2 at the bottom and Cluster 1 at the top. In fact, stations of Cluster 1 have the highest Cv and LCv value, whereas those in Cluster 2 have the lowest Cv and LCv value. Similarly, the stations of the similar region also form a group and specifies that they are homogeneous in nature.



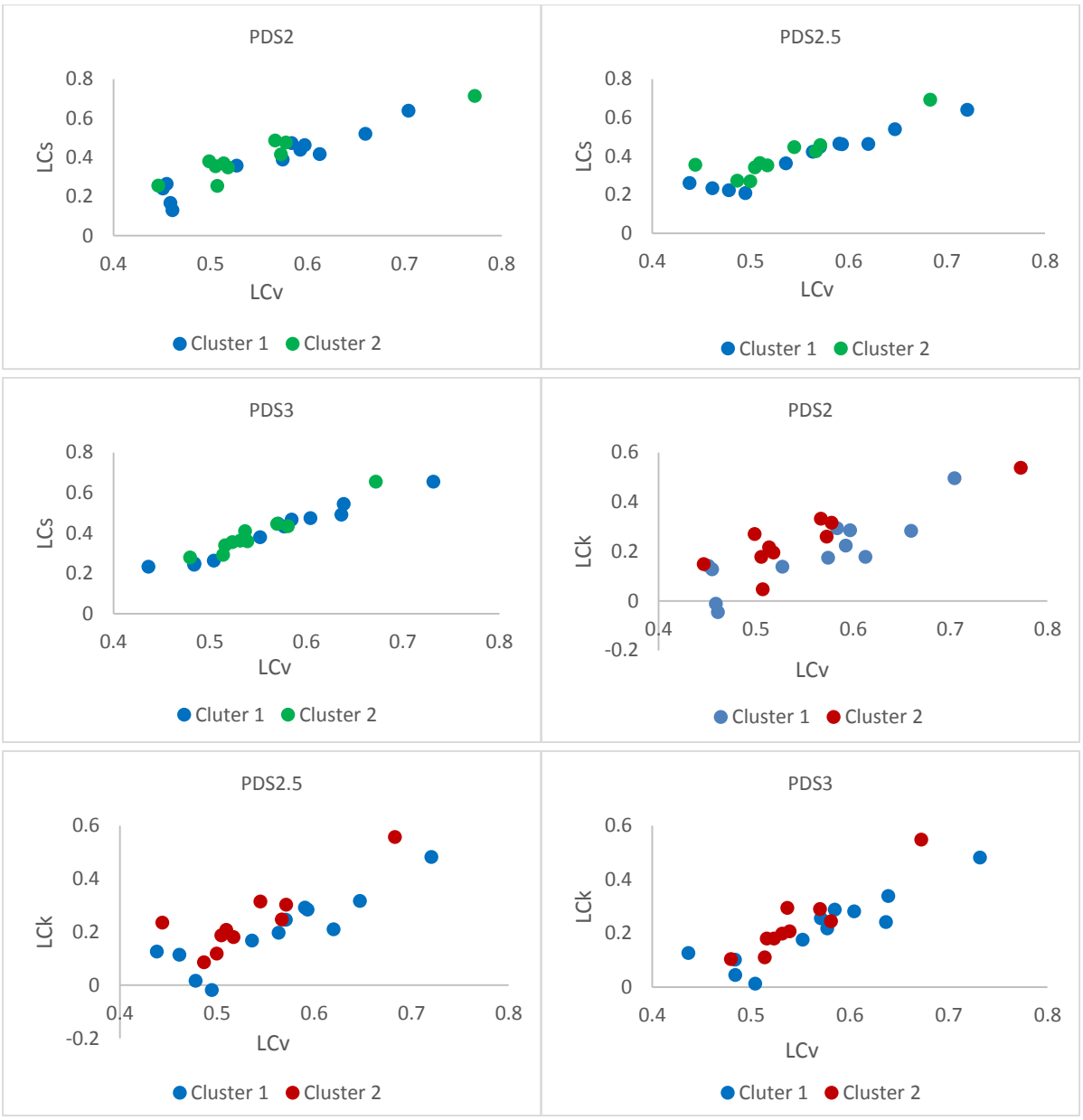
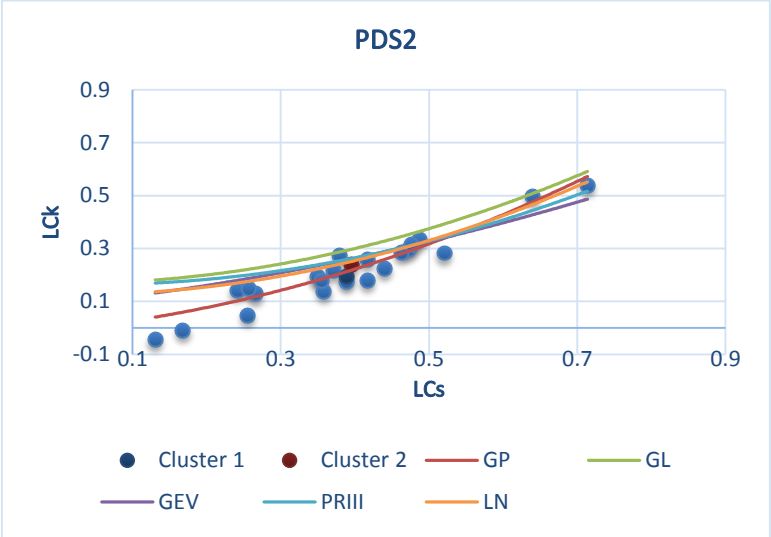


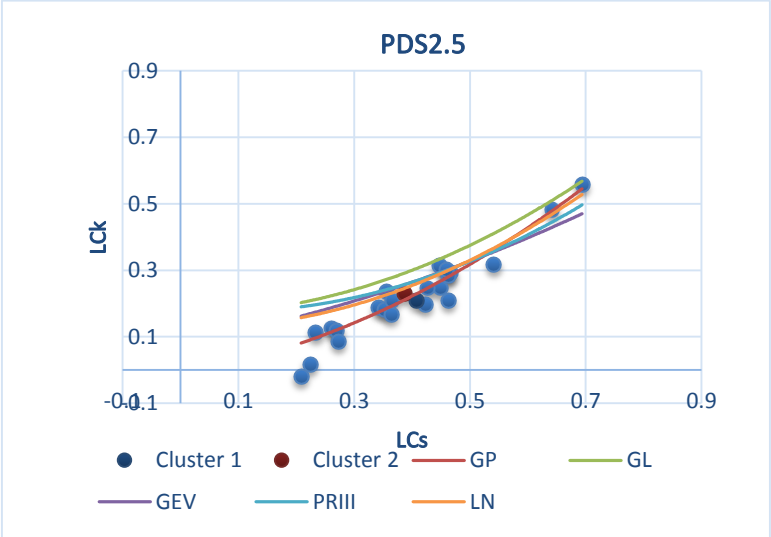
Figure 5.5: Computed values of Cv, LCv, LCv versus LCs and LCv versus LCk for PDS within the regions

5.3.5 Selection of Regional Distribution

In this part, modeling of regional frequency analysis is done, which best defines the observed data placed within a statistically homogenous region. As confirmed in the previous study of at-site analysis, the GP frequency distribution is confirmed to describe best for extracted PDS. Hence, to examine the suitability of the GP distribution for the region four other challenging distributions, namely, Generalised Extreme Value (GEV), Log Normal (LN), Pearson Type III (P3), and Generalised Logistic (GL) along with the Generalized Pareto (GP) distribution is

tested. L-moment ratio diagram and Z-statistic are performed for the region. L-moment ratio diagram (LMRD) (LCs versus LCk) is used to identify the appropriate distribution for each region (Pearson, 1993). The first step for the analysis of regionalization process, the regional L-moment coefficients are calculated using equation (5.13). Figure 5.6 shows the regional average moment ratios τ_3, and τ_4 analogues to LCs, and LCk, respectively together with a theoretical τ_3-τ_4 for the GP distribution. This also checks which PDS series is consistently best. From the figure it is shown that in all the cases of PDS, the sample τ_3 and τ_4 value lies approximately on the theoretical τ_3-τ_4 curve of the GP distribution. The Z-statistic show the inconsistency between observed and model values. It measure the excellence of fit based on the average regional sample of LCs and LCk for the fitted distributions. The regional parameters observed from L-moment algorithm for the GP distribution are given in Table 5.7.





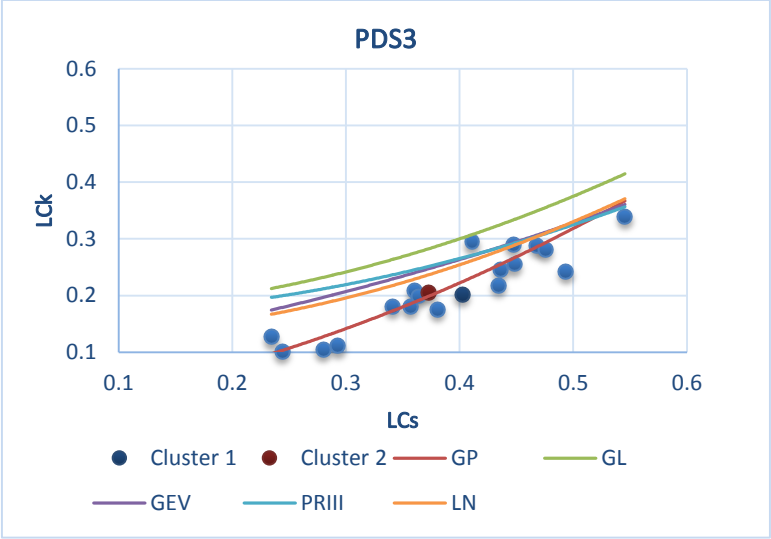


Figure 5.6: Regional LMRD for PDS2, PDS2.5 and PDS3

Table 5.7: Regional Parameters of Different Clusters

Regional parameters distribution	**Cluster 1 GP**			**Cluster 2 GP**		
	PDS2	PDS2.5	PDS3	PDS2	PDS2.5	PDS3
Shape (ζ)	-0.08046	-0.12214	-0.14749	-0.0981	-0.08571	-0.086
Scale (σ)	0.9644	0.90243	0.87403	0.90033	0.90646	0.93686
Location (μ)	-0.04878	-0.02799	-0.02525	0.00173	0.00855	-0.02543

The result found from the Z-statistic test is shown in Table 5.8, which shows the Z-values of GP distribution are within -1.64 to +1.64. This confirms that the GP is a suitable distribution in combining with extracted PDS. Thus, the GP distribution is confirmed as the best fit for regional analysis.

Table 5.8: Selected Distributions for the Regions Using Z-Statistic

GOODNESS OF FIT	**DISTRIBUTION**	**CLUSTER 1**			**CLUSTER 2**		
Z-STATISTIC	GP	PDS2	PDS2.5	PDS3	PDS2	PDS2.5	PDS3
		0.531955	0.5176594	0.5488726	0.9522931	0.9578956	0.9464553

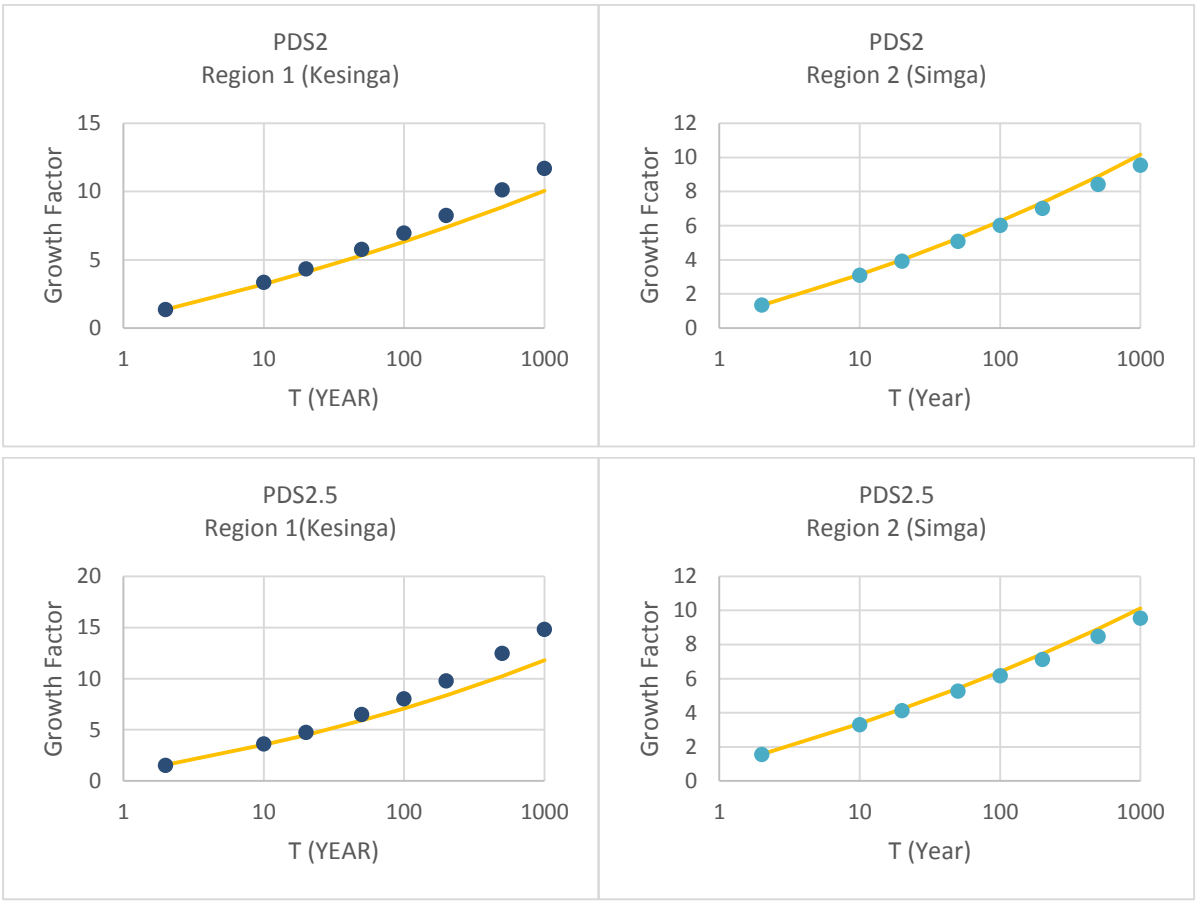
5.3.6 Derivation of Regional Frequency Curve for the Homogeneous Regions

Agreeing to the test results described above GP distributions is chosen as the best distribution

to define the partial duration series and predict satisfactory flow estimates of both regions. The regional flow frequency curve is developed by using: (i) relationship between index flood and return period, (ii) relationship between index flood and catchment characteristics.

5.3.6.1 Relationship between index flood and return period T

For different return periods, the flow quantiles are simulated using equation (5.17) and compared with the observe flow at-site station, (Figure 5.7). Finally, the regional flood frequency analysis curve evaluates the standardized flow variations for different return periods (Figure 5.8). The derived equations are used further to compute the quantiles directly of any site of interest within the regions (Table 5.9).



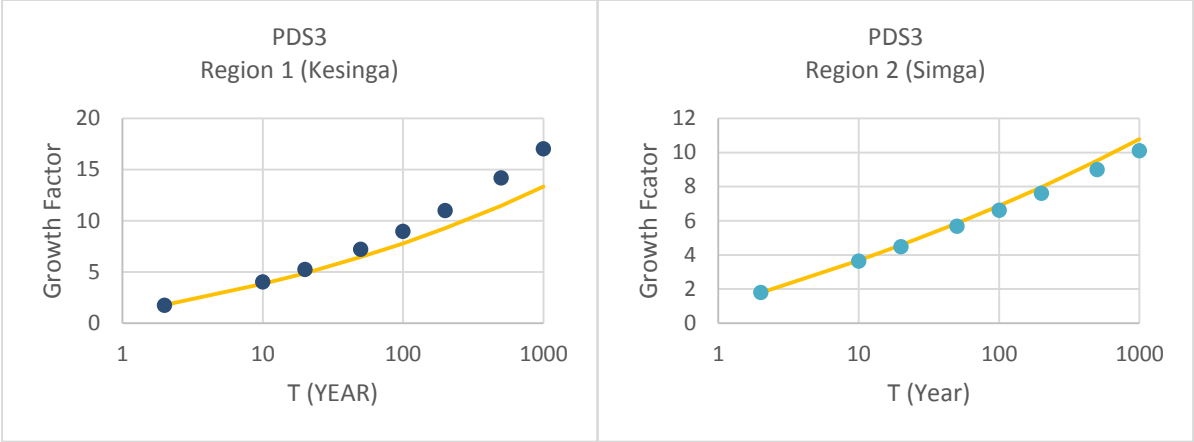


Figure 5.7: Comparison of estimated quantile of GP model using at-site (Kesinga, Region 1 and Simga, Region 2) (point) and regional (line) for PDS2, PDS2.5, and PDS3

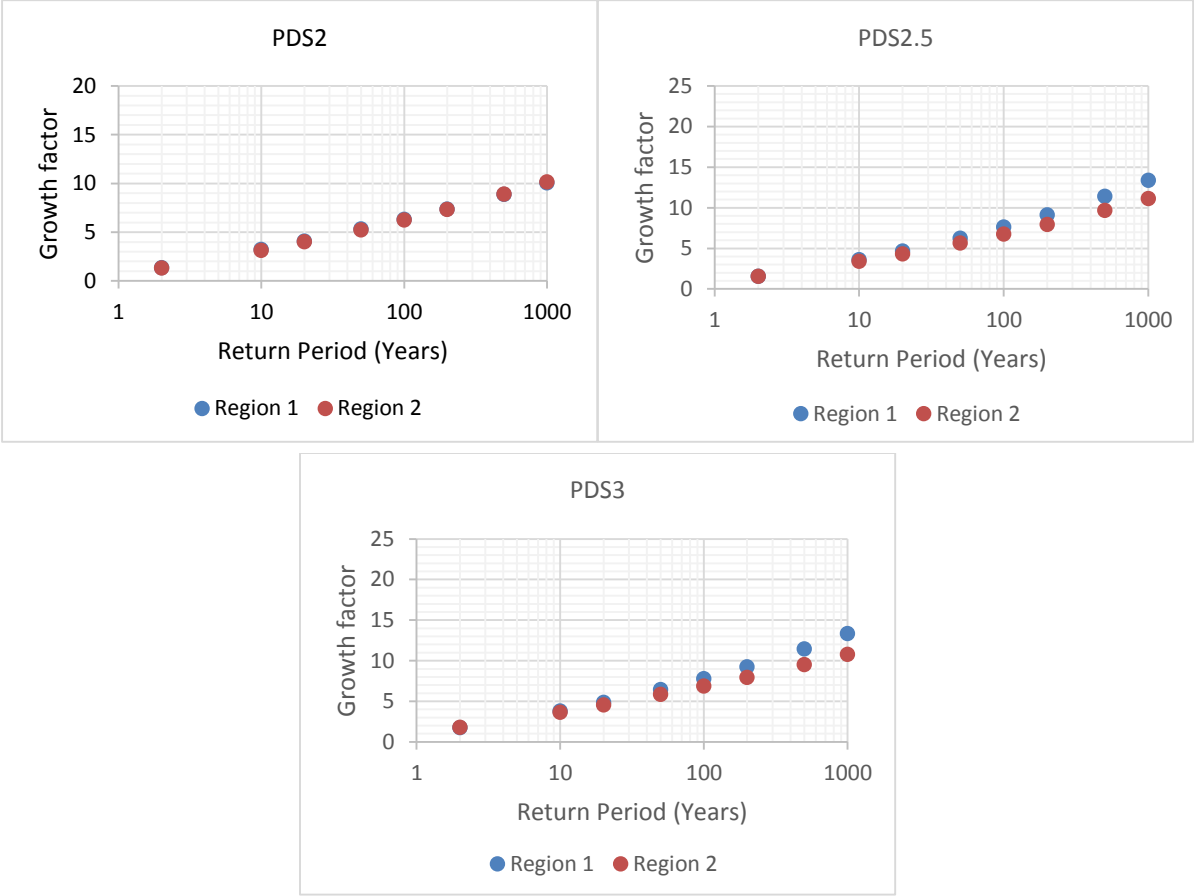


Figure 5.8: Regional flood frequency curves for two homogenous regions

Table 5.9: Standardized Quantile Values of the Region

Return Period T (Years)	Cluster 1			Cluster 2		
	PDS2	PDS2.5	PDS3	PDS2	PDS2.5	PDS3
2	$1.36 \times \overline{Q}$	$1.57 \times \overline{Q}$	$1.76 \times \overline{Q}$	$1.33 \times \overline{Q}$	$1.57 \times \overline{Q}$	$1.79 \times \overline{Q}$
10	$3.21 \times \overline{Q}$	$3.53 \times \overline{Q}$	$3.83 \times \overline{Q}$	$3.13 \times \overline{Q}$	$3.36 \times \overline{Q}$	$3.67 \times \overline{Q}$
20	$4.09 \times \overline{Q}$	$4.49 \times \overline{Q}$	$4.88 \times \overline{Q}$	$4.00 \times \overline{Q}$	$4.22 \times \overline{Q}$	$4.57 \times \overline{Q}$
50	$5.32 \times \overline{Q}$	$5.90 \times \overline{Q}$	$6.45 \times \overline{Q}$	$5.24 \times \overline{Q}$	$5.42 \times \overline{Q}$	$5.84 \times \overline{Q}$
100	$6.32 \times \overline{Q}$	$7.08 \times \overline{Q}$	$7.79 \times \overline{Q}$	$6.25 \times \overline{Q}$	$6.40 \times \overline{Q}$	$6.88 \times \overline{Q}$
200	$7.37 \times \overline{Q}$	$8.36 \times \overline{Q}$	$9.27 \times \overline{Q}$	$7.34 \times \overline{Q}$	$7.44 \times \overline{Q}$	$7.97 \times \overline{Q}$
500	$8.86 \times \overline{Q}$	$10.23 \times \overline{Q}$	$11.47 \times \overline{Q}$	$8.89 \times \overline{Q}$	$8.92 \times \overline{Q}$	$9.52 \times \overline{Q}$
1000	$10.05 \times \overline{Q}$	$11.79 \times \overline{Q}$	$13.35 \times \overline{Q}$	$10.16 \times \overline{Q}$	$10.11 \times \overline{Q}$	$10.78 \times \overline{Q}$

Table 5.9 indicates that, the quantile values obtained for all PDS for both at-site and region for the return period less than or equal to 100 years gives good results in comparison to the results obtained using the return period T> 100. This may result in an underestimation or overestimation of 1000 year-quantiles at some stations when it works a regional distribution.

5.3.6.2 Relationship between index flood and catchment characteristics

In this study regression analysis was applied to predict mean partial duration flood for ungauged catchments. Nine variables are selected as per factor analysis for clustering, and then the regression has been accomplished by taking these variables. Initially, a preliminary correlation matrix, which consists of coefficients of correlation, was conducted between pairs of catchment characteristics and index flood for both region in the study area. For each region, nine independent variables were applied to select the best regression model by considering the computed value of the correlation calculation. The resulting regression equations to predict the index flood for both clusters in power form are as follows:

FOR CLUSTER 1

$$\overline{Q}_{PDS2} = CA^{3.02} * SL^{-0.78} * DD^{-0.22} * CR^{1.68} * PBF^{0.41} * AMP^{0.34} * PBA^{0.17} * LSL^{-2.89} * PBB^{-0.19}$$

$$\overline{Q}_{PDS2.5} = CA^{2.98} * SL^{-0.75} * DD^{-0.27} * CR^{1.63} * PBF^{0.51} * AMP^{0.33} * PBA^{0.21} * LSL^{-2.77} * PBB^{-0.18}$$

$$\overline{Q}_{PDS3} = CA^{2.93} * SL^{-0.73} * DD^{-0.33} * CR^{1.62} * PBF^{0.60} * AMP^{0.32} * PBA^{0.25} * LSL^{-2.67} * PBB^{-0.17}$$

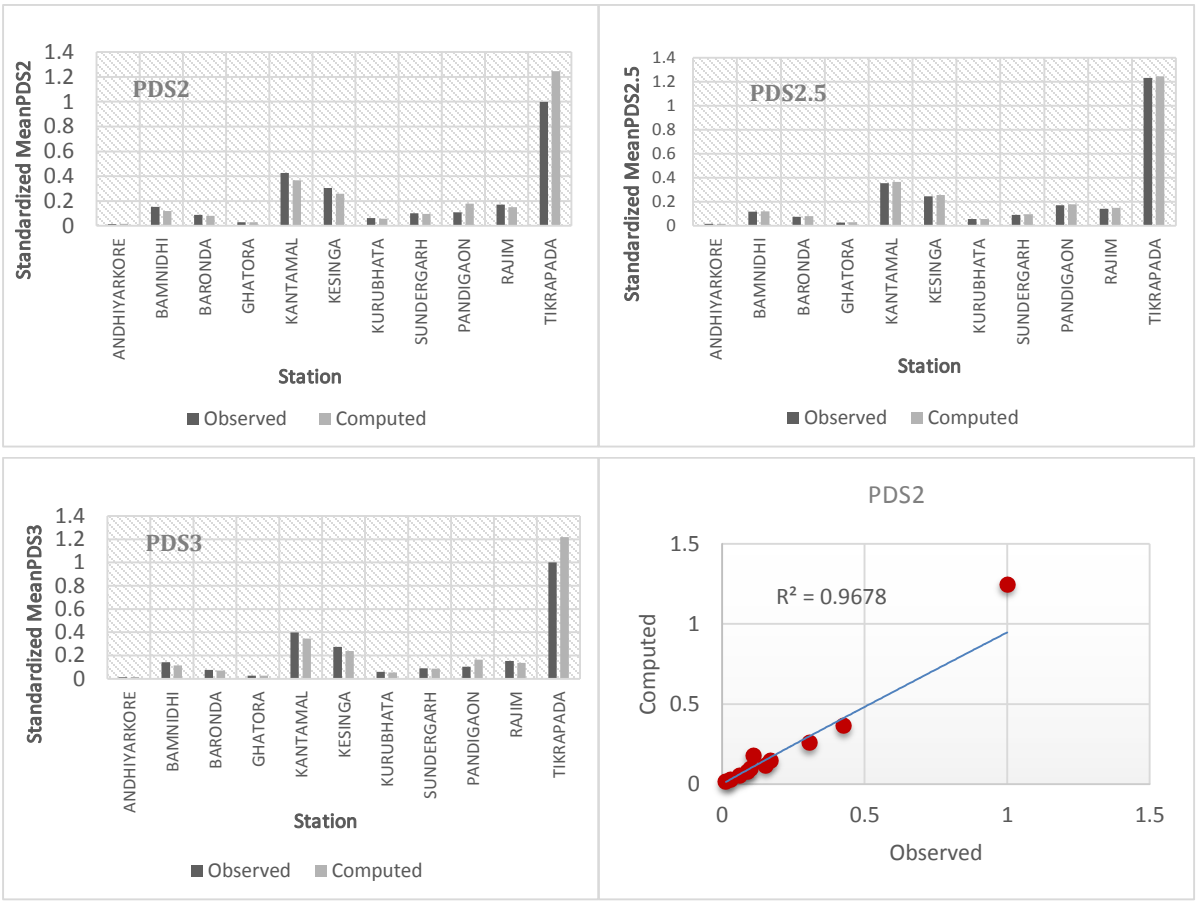
FOR CLUSTER 2

$$\overline{Q}_{PDS2} = CA^{0.95} * SL^{-2.33} * DD^{-3.92} * CR^{5.22} * PBF^{5.88} * AMP^{-8.08} * PBA^{18.43} * LSL^{-7.69} * PBB^{0.95}$$

$$\overline{Q}_{PDS2.5} = CA^{1.24} * SL^{-1.44} * DD^{-3.33} * CR^{6.05} * PBF^{3.88} * AMP^{-7.08} * PBA^{13.93} * LSL^{-6.63} * PBB^{0.78}$$

$$\overline{Q}_{PDS3} = CA^{1.21} * SL^{-1.65} * DD^{-3.46} * CR^{6.05} * PBF^{4.32} * AMP^{-7.46} * PBA^{14.99} * LSL^{-6.99} * PBB^{0.82}$$

The performance of these equations are validated for all gauged sites of each region by assuming the basin as an ungauged basin (Figure 5.9 and Figure 5.10).



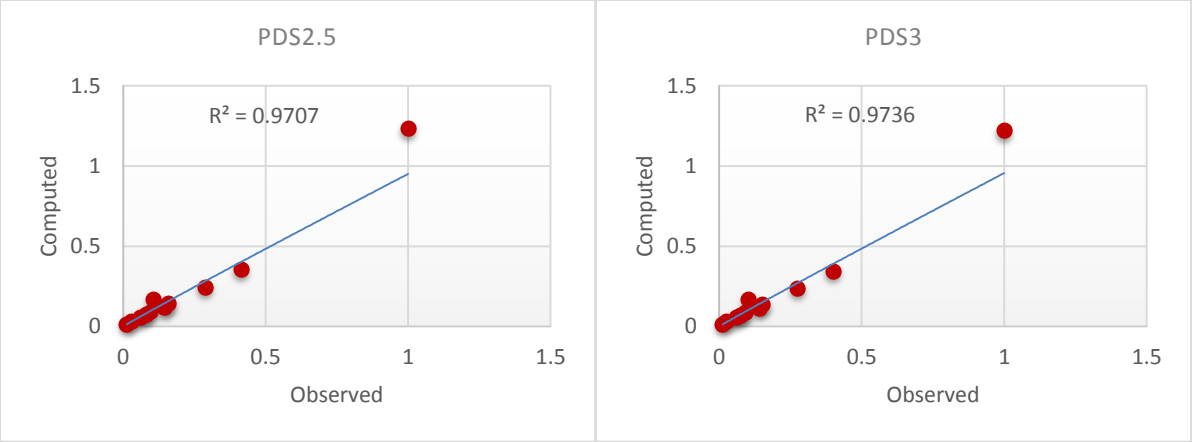
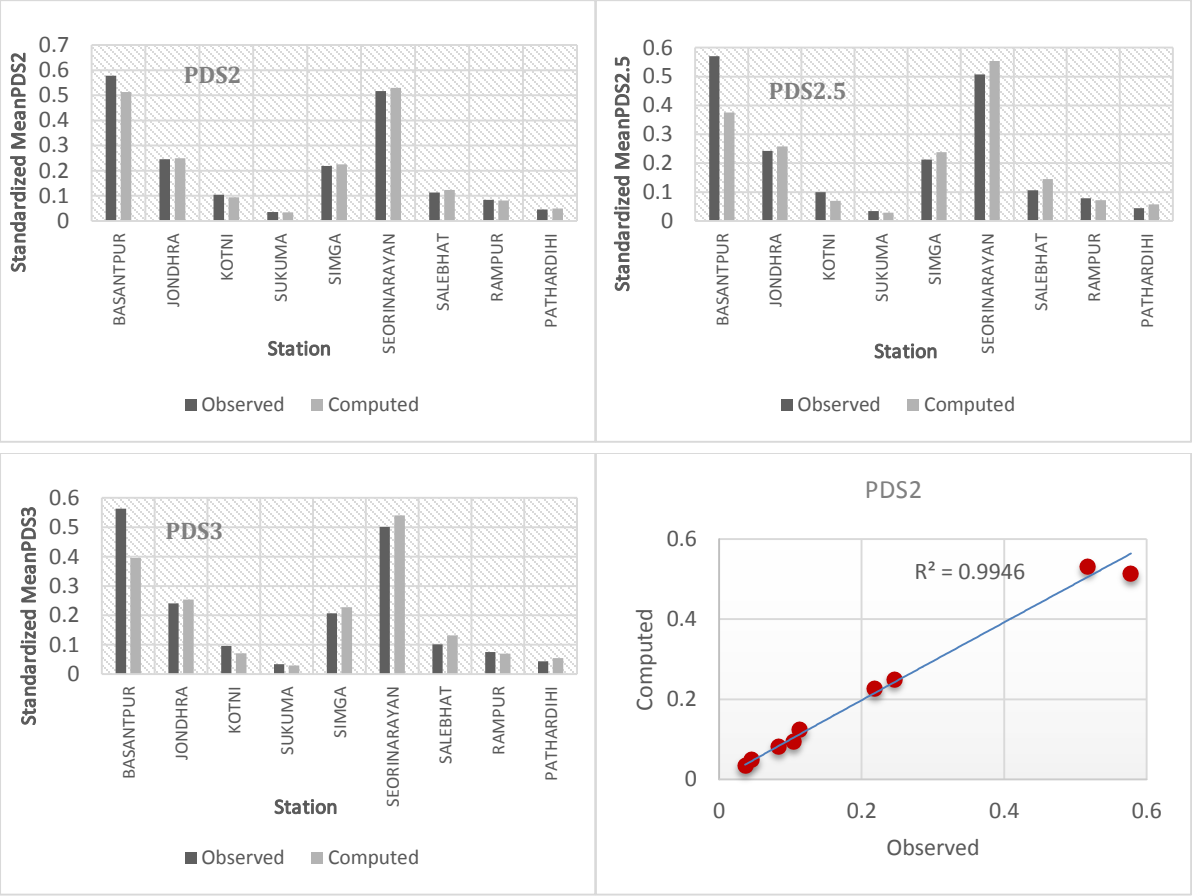


Figure 5.9: Comparison of observed and estimated index flood based on regression model for Region 1



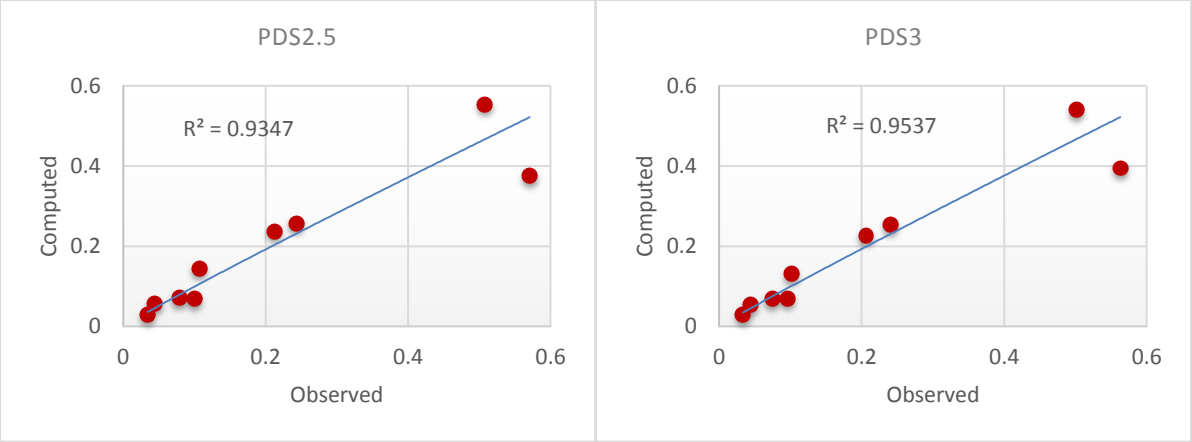


Figure 5.10: Comparison of observed and estimated index flood based on regression model for Region 2

Figures 5.9 and 5.10 shows that the predicted and observed index floods of PDS2, PDS2.5 and PDS3 have a strong correlation with correlation coefficients varies from 0.93 to 0.97.

5.4 SUMMARY

The Mahanadi river basin is divided into two homogeneous regions. The KM clustering method remains robust and consistent for different variables. The GP distribution was found to be suitable for both the regions. For both regions, regional frequency curves for PDS2, PDS2.5 and PDS3 are calculated and found that the values in case of PDS2.5 and PDS3 increase for higher return periods. The quantiles of ungauged sites within each region are estimated. A minor difference is observed between estimated quantiles using the on-site and regional distribution. It is found more consistent for return periods less than or equal to 100 years ($T \leq 100$ years). Regarding the index flood estimation, i.e., the correlation between the observed and predicted flood, and the derived regression equations made encouraging results for both the regions. In conclusion, regional flood frequency analysis delivers vigorous evidence for the planning, design, and operation of hydraulic structures, flood management, and mitigation measures.

CHAPTER 6
APPLICATION OF SOFT COMPUTING TECHNIQUES FOR RIVER FLOW PREDICTION IN LOWER MAHANADI RIVER BASIN USING PARTIAL DURATION SERIES

6.1 INTRODUCTION

Flood is a regular phenomenon in lower Mahanadi river basin due to inadequate water holding capacity of the river channel and low retention capacity of the soil of the floodplain area during monsoon. Any discharge above this creates a flood like situation and inundates the densely populated area causing immense loss to life and property. The discharge released from Hirakud dam during the monsoon in addition to the heavy rainfall at the downstream site contributes largely to the flood in the region. It is found essential, to carry out a detailed study and establish a workable flood forecasting method for lower Mahanadi basin (Figure 6.1).

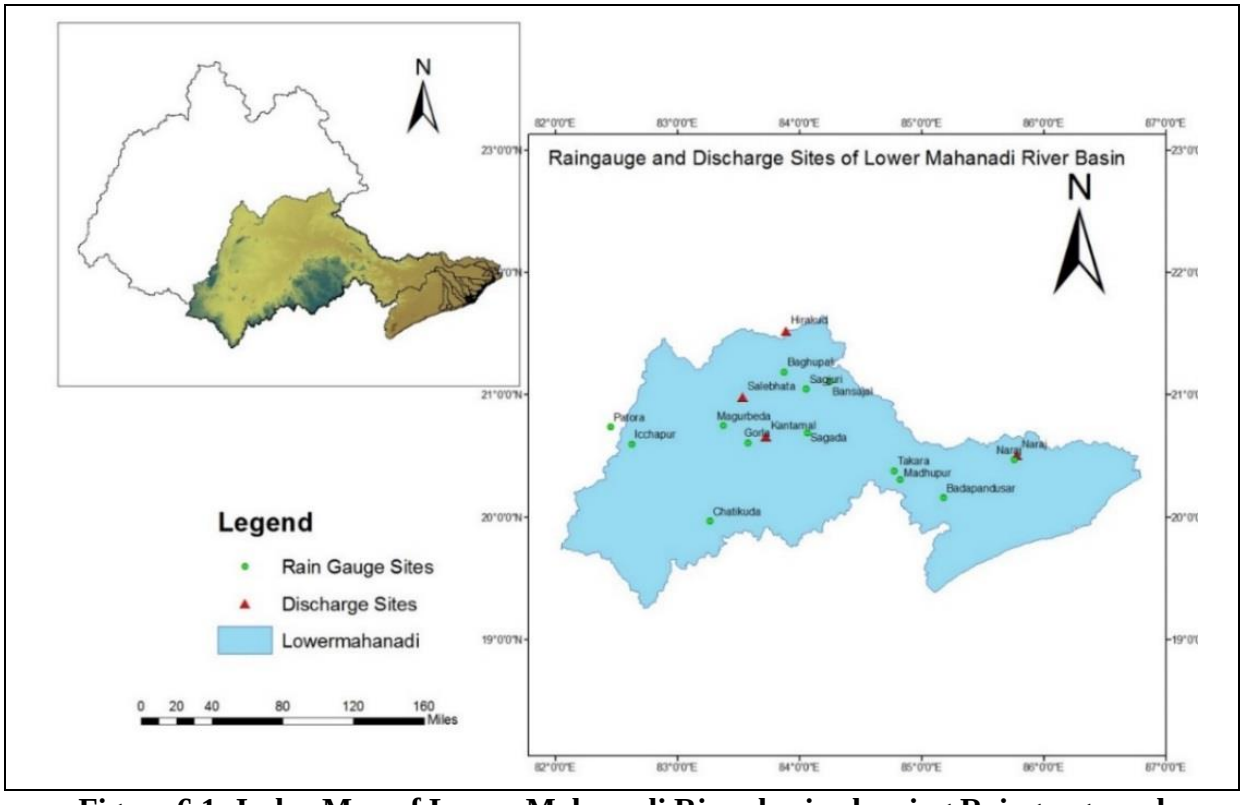


Figure 6.1: Index Map of Lower Mahanadi River basin showing Raingauge and Discharge Sites

Keeping above in view, the present study emphasizes on, (i) The trends analysis of different peak discharge of partial duration series having an average number of flood peaks per year (i.e

λ=3) at Naraj, Hirakud, Kantamal and Salebhata for the period of 2000–2009 using both the Mann–Kendall test and linear regression test. (ii) The trend analysis of extremes monthly maximum areal rainfall computed from Thiessen polygon method (Figure 6.2) in the river basin for the period of 1901–2010 and also, mean areal rainfall of partial duration series having an average number of flood peaks per year (i.e., λ=3) using Thiessen polygon method (Figure 6.3) of lower regions. (iii) Generating random values with a best-fit probability distribution for prediction of discharge of higher return periods, and (iv) Application of ANN and ANFIS for prediction of the peak runoff at Naraj station.

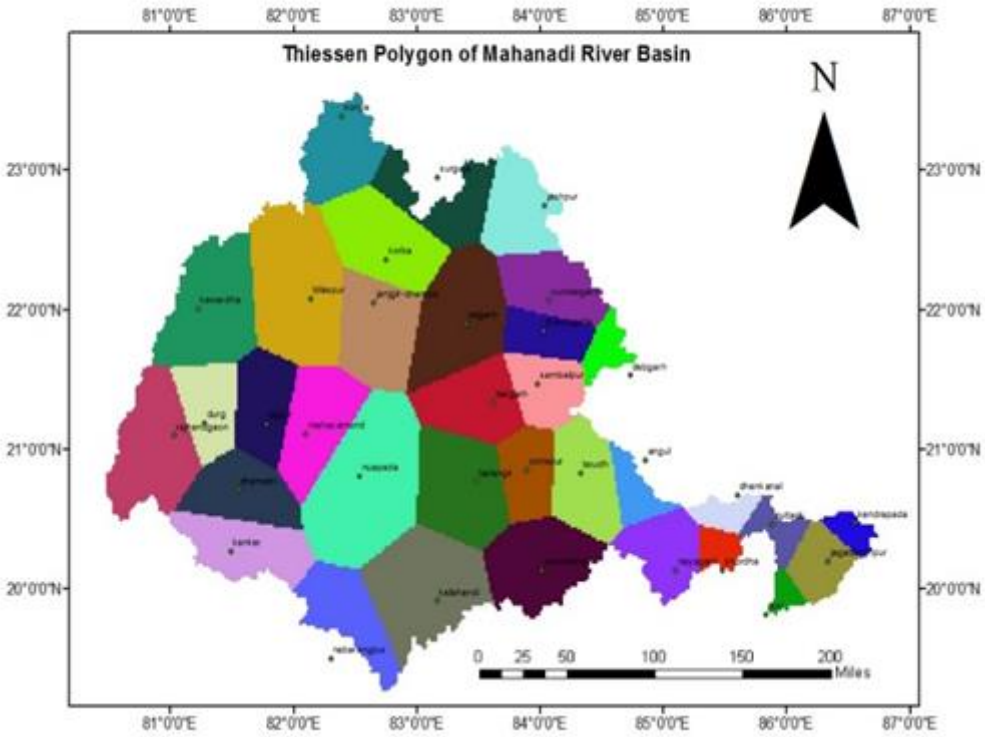


Figure 6.2: Monthly maximum areal rainfall using Thiessen polygon method in upper and lower regions

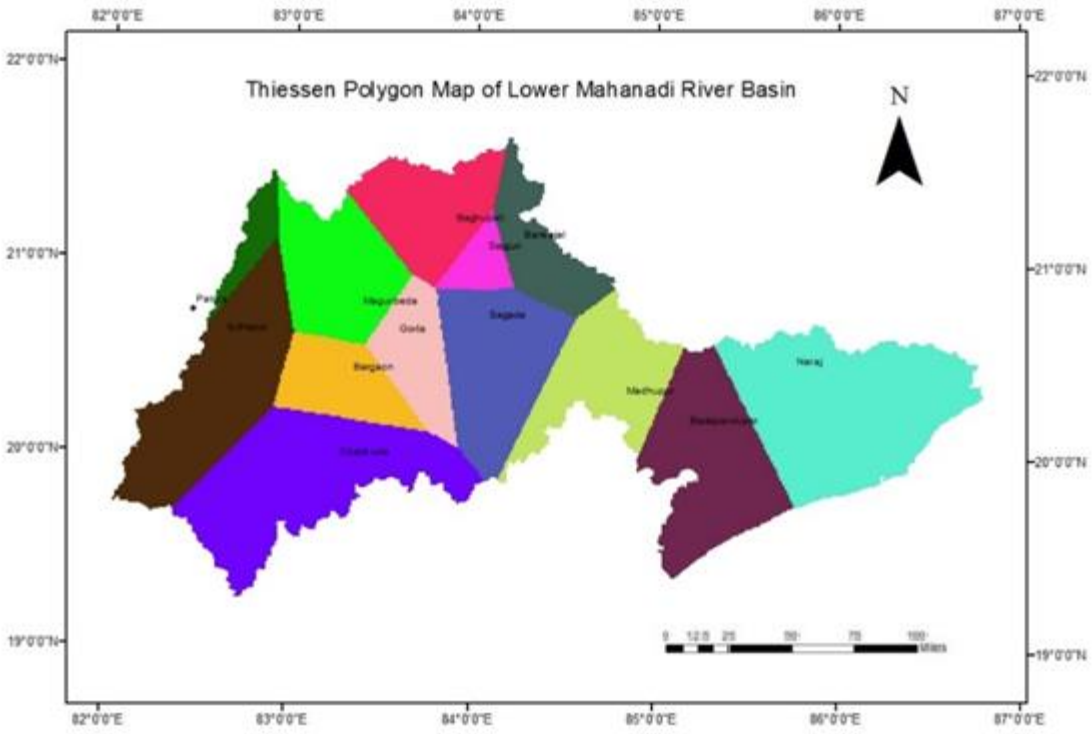


Figure 6.3: Mean areal rainfall using Thiessen polygon method in lower regions

The study was carried out in the lower Mahanadi River basin shown in Figure 6.1, which begins at Hirakud Dam and ends at the Bay of Bengal. Discharge of threshold value 25250 m^3/s of partial duration series having an average number of flood peaks per year (i.e., λ=3) and above at Naraj stations are considered because an existing embankment system in the Mahanadi delta mitigates floods up to 28,400 m^3/s. Also, extreme monthly maximum mean areal rainfall in upper and lower regions of the river basin for the period of 1901–2010, and mean areal rainfall which is extracted from partial duration series having an average number of flood peaks per year (i.e., λ=3) using Thiessen polygon method in lower region analysis for the period 2000–2009 are used. A total of 30 peaks discharge datasets for 10 years at each sites were obtained and 300 peak discharge datasets are generated for each site using the best fit probability distribution. The datasets are used as the input variable for further analysis. Out of this 300 peaks, 70% peak values are considered for training and 30% peak values are considered for testing of models. Discharge data (Q in m^3/s) was converted into runoff data (in mm/d) using the relation (Q*86.4/A), where A is the catchment area (in km^2) of the station. Figure 6.2 shows the Thiessen polygon for entire basin and lower basin.

6.2 METHODOLOGY

6.2.1 Trend analysis

The trend is generally known as a steady increase or decrease of the time series characteristics caused due to natural or man-made changes like deforestation, urbanization, large-scale landslide, large changes in watershed conditions. In the present study, most commonly used Mann-Kendall test and linear regression test are applied for detecting and predicting trends in lower Mahanadi river basin.

Trend detection using Mann-Kendall test is discussed by many researchers in the past (Goswami et al., 2006; Dufek and Ambrizzi, 2008; Wan Zin et al., 2010; Pal and Al-Tabbaa, 2011a, b; Douglas and Chelsea, 2000). In fact, the Mann-Kendall S Statistic is computed as:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign\left(T_j - T_i\right) \tag{6.1}$$

$$sign\left(T_j - T_i\right) = \begin{cases} 1 if T_j - T_i > 0 \\ 0 if T_j - T_i = 0 \\ -1 if T_j - T_i < 0 \end{cases}$$

where Tj and Ti are the annual values in years j and i, j > i, respectively.

If n < 10, the value of |S| is compared directly to the theoretical distribution of S derived by Mann-Kendall and the two-tailed test is used. At certain probability level H_0 is rejected in favor of H_1 if the absolute value of S equals or exceeds a specified value $S_{\alpha/2}$, where $S_{\alpha/2}$ is the smallest S which has the probability less than $\alpha/2$ to appear in case of no trend. A positive (negative) value of S indicates an upward (downward) trend. For $n \geq 10$, the statistic S is approximately normally distributed with the mean and variance as follows:

$$E(S) = 0 \tag{6.2}$$

The variance (σ^2) for the S-statistic is defined by:

$$\sigma^2 = \frac{n(n-1)(2n+5) - \sum t_i (i)(i-1)(2i+5)}{18} \tag{6.3}$$

where t_i denotes the number of ties to extent i. The summation term in the numerator is used only if the data series contains tied values. The standard test statistic Z_s is calculated as follows:

$$Z_s = \begin{cases} \frac{s-1}{\sigma} \; for s > 0 \\ 0 \; for s = 0 \\ \frac{s+1}{\sigma} \; for s < 0 \end{cases} \quad (6.4)$$

The test statistic Z_s is used as a measure of the significance of the trend. In fact, this test statistic is used to test the null hypothesis, H0. If $|Z_s|$ is greater than $Z_{\alpha/2}$, where α represents the chosen significance level (eg: 5% with Z 0.025 = 1.96) then the null hypothesis is invalid implying that the trend is significant.

Also, a simple linear regression technique has been used to detect the trend. It shows the positive or negative slope values which denote the direction and magnitude of the trend. A regression line of the order y=a+bx is fitted to the sample data. If b is found significantly different from zero, then we assume trend to be present in the data. The superiority of these linear regression based trends is analyzed using the r^2 values.

6.2.2 Random Number Generation

A random number generation (RNG) is a computational or physical analysis to generate a sequence of numbers that lack any pattern, i.e., appear random. The many applications of randomness have led to the development of several different methods for generating random data. The high-quality random numbers have been generated using Mersenne Twister algorithm, a probability distributions function, in EasyFit software. The generator has a period of 2^{19937}-1 (more than 10^{6000}) and passes numerous tests for statistical randomness, including the well-known Diehard tests (a number of statistical tests for measuring the quality of a set of random numbers). These qualities, along with its high speed, make the Mersenne Twister generator an algorithm of choice for most statistical simulations.

6.2.3 Development of Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) Models

The basic concepts that comprise the neural network approach and neuro-fuzzy theory such as weights, learning algorithm, fuzzy set, membership functions, the domain partitions, and fuzzy if–then inference rules are described in different research papers and textbooks (Goldberg 1989; Dawson and Wilby 1998). A brief description is given below.

6.2.3.1 ANN Analysis

The ANN is a very powerful computational algorithm, which is used to simulate complex nonlinear relationships, especially in situations where the explicit form of the relation between the variables involved is unknown. Neural networks are composed of interconnected parallel structure. It generally consists of three layers, where data are provided to the network of ANN, the data are processed in hidden layers, and the results of the input layer are produced in the output layer. A neural network consists of weights which help to improve the training process by adjusting the connecting weights between the elements. Usually, two types of neural network are used in studying many hydrological problems, such as multilayer perceptron (MLP) and radial basis functions (RBF). We used MLP neural network using STATISTICA Software for the forecasting over the RBF because it is mostly used the neural network with a high number of applications and capable of modeling complex functions by ignoring irrelevant inputs and noise. The details are described below.

Multilayer perceptron (MLP)

MLP is a layered network having the input, hidden and output layers, denoted as the perceptron. It also consists of nonlinear processing elements (PE) which are weighted by a scalar weight (*w*) and bias (b) as shown in Figure 6.4. The processing elements produce the final output from the inputs, using the nonlinear activation functions. It is a trial-and-error process for providing excellent results and basically depends on the data availability and problem type.

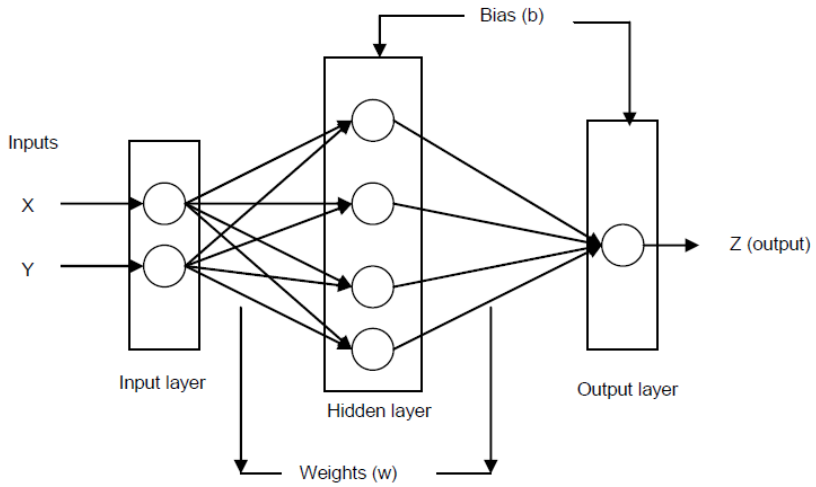


Figure 6.4: General Structure of MLP

During the training process, the weights are adjusted to minimize the error between the output and the PE. The optimal weights are the product of the inverse of the input autocorrelation

matrix (R^{-1}) and the cross-correlation vector (P) between the input and the desired response. The analytical solution of this problem is equivalent to a search technique to obtain the minimum of the quadratic performance surface, $J(wi)$, using gradient descent by adjusting the weights at each epoch (Haykin, 1999):

$$w_i(k+1) = w_i(k) - \eta \nabla J(k) \qquad \nabla J_i = \frac{\partial J}{\partial w_i} \tag{6.5}$$

where η is the learning rate coefficient and $\nabla J_i(k)$ is the gradient vector of the performance surface at iteration (k) for the i^{th} input node. Equation (2) is used to calculate the performance surface (J):

$$J = \sum_p (d_p - y_p)^2 \qquad \text{and} \quad \min J \rightarrow w_{opt} = R^{-1}P \tag{6.6}$$

where w_{opt} is the optimal weight, d_p is the target output, and y_p is the computed output of the p^{th} output neuron.

MLP use the back-propagation algorithm (Fausett, 1994; Patterson, 1996; Haykin, 1999) because it is easy to recognize and can be effectively used in many applications. Three back-propagation training algorithms were used to train the models such as the Levenberg-Marquardt (LM), the gradient descent algorithm with variable learning and momentum factor (GDX), and the conjugate descent algorithm (CGF). Then the performance of all above algorithms was compared based on the root mean squared error (RMSE), Nash-Sutcliffe efficiency (E), the index of agreement (d) and correlation coefficient (R^2).

The Levenberg-Marquardt (LM) algorithm uses a second-order training mode without computing Hessian matrix (Demuth and Beale, 1998). When the performance function has the form of sum of squares, the Hessian matrix can be estimated as $H = J^T J$ and the gradient can be calculated as $g = J^T \boldsymbol{e}$, where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and $\boldsymbol{e}$ is a vector of network errors. The LM algorithm uses the above approximation update as $w_{k|+1} = w_k - [H + \mu I]^{-1} J^T e$, where w indicates the weight of the neural network, and μ is a non-negative scalar that controls the learning process. The LM method is the standard method for minimization of the MSE criterion, due to its rapid convergence properties and robustness.

The gradient descent with momentum and adaptive learning rate algorithm (GDX) can train any network. It calculates derivatives of performance (*p)* with respect to the weight and bias variables (X) (Hagan et al., 1996). Each variable is adjusted by: $dX = mc\ dX_{prev} + lr\ mc\ (\partial p/\partial x)$, according to gradient descent with momentum. Where, dX_{prev} is the previous change in the weight or bias, mc is the momentum coefficient, lr is the learning rate, dX is the weight correction, and the partial derivative ($\partial p/\partial x$) is the performance change with respect to change in the weight. If the performance decreases, then the learning rate is increased for each epoch. A detailed description of this training method is given in (Hagan et al., 1996).

The conjugate gradient back-propagation algorithm (CGF) can also train any network. Back propagation is used to calculate derivatives of performance p with respect to the weight and bias variables X. Each variable is adjusted according to the expression $X = X + adX$, where dX is the search direction. Where, a is a constant which minimizes the error functions along the search direction. The first search direction is negative of the gradient of performance. In following epochs, the search direction is calculated from the new gradient and the previous search direction according to the relationship: $= dX = -G_X + dX_{old}Z$, where G_X is the gradient. The parameter Z can be computed in several different ways. For the Fletcher-Reeves variation of conjugate gradient, Z is computed according to $(GX_{new})^2 / (GX_{prev})^2$ where GX_{prev} is the previous gradient, and GX_{new} is the current gradient. A detailed description of the conjugate gradient algorithm is given in (Fletcher, 1987).

Here, different steps for ANN to solve the flood forecasting problems is described. At first normalization and scaling is done. Secondly, an ANN network architecture is fixed where the number of hidden layers, neurons in each layer and the connectivity between neurons are set. Many experimental results say that one hidden layer may be enough for most forecasting problems. The number of neurons in each layer depends upon the problem being studied. Less number of neurons in hidden layer will make the network with less degree of freedom for learning, and more number of neurons will lead towards more time and over fitting. Validation set error is often used to determine the optimal number of hidden neurons for a given study. In the third step is the finalization of a learning algorithm for training the network. The parameters are finalized for the training data set to be applicable for any kind of testing data. ANN architecture is considered to be trained when the difference between ANN output and observed output is very small. Finally, the validation step is applied to get the performance of the network. The optimal number of hidden neurons and training iterations can also be determined through validation. The selection of acceptable model is finalized on the basis of RMSE, R^2, d and E.

6.2.3.2 ANFIS Modeling

The adaptive neuro-fuzzy inference system (ANFIS) is a soft computing method which combines the feature of both ANN and fuzzy inference system (FIS). ANN has the capability of self-learning and self-adapting the data for forecasting but difficult to understand the learning process. However, the fuzzy logic models are easy to implements a nonlinear mapping which is skilled by a number of fuzzy IF–THEN rules to define the local performance of mapping. The fuzzy membership parameters are optimized either by using a back-propagation algorithm or by a combination of both back-propagation and least square method, and their efficiency depends on the estimated parameters. ANFIS model was first used symmetrically by (Takagi and Sugeno, 1985), and they found numerous applications in the field of the prediction.

The structure and parameter adjustment

The structure of the ANFIS is similar to that of a neural network. An Adaptive Neuro-Fuzzy Inference System consisted of five important functional building parts of the fuzzy logic toolbox, those are (i) rule base, (ii) database, (iii) decision-making unit, (iv) fuzzification interface and (v) defuzzification interface shown in Figure 6.5.

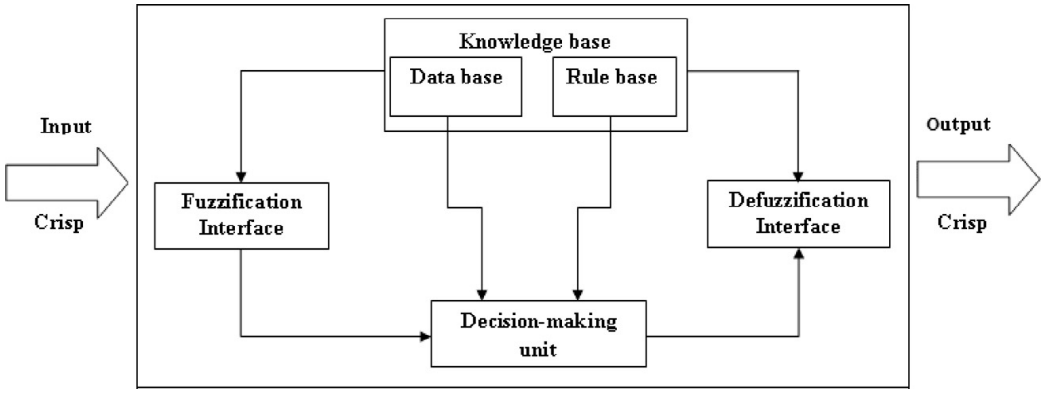


Figure 6.5: Block diagram of fuzzy based inference system

To explain the working principle of ANFIS an example is taken, consider that x and y are the two inputs and z is the output. The first-order IF–THEN fuzzy rules can be expressed as follows:

Rule 1: If x is A_1 and y is B_1 then $f_1 = p_1x + q_1y + r_1$

Rule 2: If x is A_2 and y is B_2 then $f_2 = p_2x + q_2y + r_2$

If f_1 and f_2 are constants instead of linear equations, we have zero order TSK fuzzy models. The node function in the same layer is of the same function family as described below.

Layer 1: Each node in this layer created a membership grade of a linguistic label. For instance, the node function of the *i*th node would be

$$O_i^j = \mu A_i(x) = \frac{1}{1+\left[\left(\frac{x-c_j}{a_i}\right)\right]^{b_i}} \tag{6.7}$$

Here, O_i^j can be denoted the output of the *i*th node in layer *j*.

where x was the input to node *i* and A_i is the linguistic label (small, large) associated with the node. The parameters that changed the shapes of the membership function are {*ai,bi,ci*}.The parameters in this layer were known as Premise parameters.

Layer 2: Each node in this layer finds the firing strength of each rule via multiplication, given as

$$O_i^2 = w_i = \mu A_i(x) \times \mu B_i(y), \qquad \text{where } i=1,2. \tag{6.8}$$

Layer 3: Here, the ith node finds the ratio of ith rule's firing strength to the sum of all rule's firing strengths as

$$O_i^3 = \overline{w_i} = \frac{w_i}{w_1 + w_2}, \quad \text{where } i=1,2. \tag{6.9}$$

The layer also called as normalised firing strengths.

Layer 4: Every node *i* in this layer is a squared node with a node function as given below:

$$O_i^4 = \overline{w_i} f_i = \overline{w_i}\left(p_i + q_{iy} + r_i\right) \tag{6.10}$$

where $\overline{w_i}$ is the parameter set as the output of layer 3.The parameters in this layer were known as Consequent parameters.

Layer 5: Here, the summation of all incoming signals is computed by the single circle node, and is given as:

$$O_i^5 = overalloutput = \sum_i^n \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{6.11}$$

The output layer backward to input nodes recursively. The back propagation learning rule used here is exactly same as in common feed forward neural network. An adaptive neural network structure presented in Figure 6.6 is functionally similar to fuzzy inference system. It is observed that the values of premise parameters and the overall output '*f*' are a linear combination of the consequent parameters. The output *f* can be formulated as

$$f = \frac{w_1}{w_1 + w_2} f_2 + \frac{w_1}{w_1 + w_2} f_2 = \overline{w} f_1 + \overline{w} f_2$$

$$= \left(\overline{w}x\right)p_1 + \left(\overline{w}y\right)q_1 + \left(\overline{w_1}\right)r_1 + \left(\overline{w_2}x\right)p_2 + \left(\overline{w_2}y\right)q_2 + \left(\overline{w_2}\right)r_2 \tag{6.12}$$

The output *f* is linear in the consequent parameters p_1, q_1, r_1, p_2, q_2, r_2.

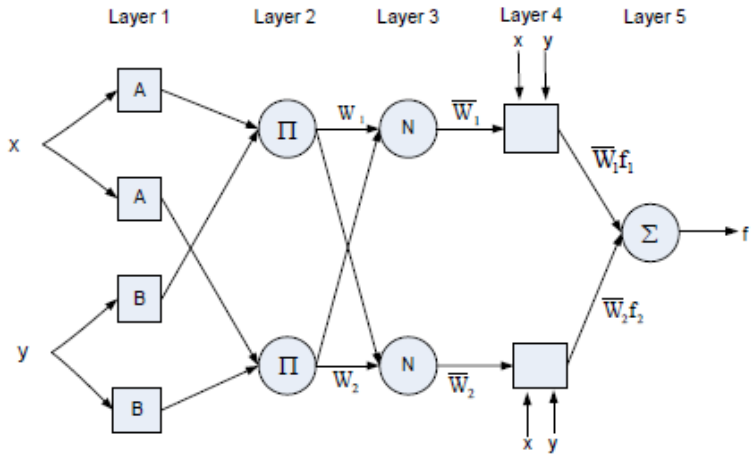


Figure 6.6: A typical architecture of ANFIS system

Again, the Hybrid learning rule combined a gradient descent and the least squares method to find a feasible of antecedent and consequent parameters. The details of the hybrid learning rule are described by (Jang et al., 1997). In this study, a hybrid learning algorithm is adopted to obtain optimal parameter of ANFIS using MATLAB 2014. The flow chart for complete approach and ANFIS algorithm is shown in Figure 6.7.

6.2.3.3 Performance Evaluation of ANN and ANFIS

The performance of both the models has been carried out using root mean squared error (RMSE) and The Nash–Sutcliffe efficiency (E). In addition, the index of agreement (d) and correlation coefficient (R^2) are also used to evaluate the performance of the models.

Root Mean Square Error (RMSE): Root Mean Squared Error or Root Mean Squared Deviation was a measure of the differences between values predicted by model or an estimator and the actually observed values. These individual differences were called as residuals when the calculations were performed on the data sample that was used for estimation and were known as estimation errors when computed out of the sample. The RMSE was served to aggregate the magnitudes of errors in predictions for various times into single measure of predictive power.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}} \tag{6.13}$$

where O_i is the observed value, P_i is the model-simulated value, and n is the total number of observations of the dataset.

The Nash–Sutcliffe Efficiency (E) (Nash and Sutcliffe, 1970), defined as

$$E = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{6.14}$$

where $O_{i,}$ is the observed value, $P_{i,}$ is the model-simulated value, and $\overline{O}$ is the mean observed value.

The Nash-Sutcliffe model efficiency is used to assess the predictive power of hydrological models. Nash-Sutcliffe efficiencies can range from -8 to 1. An efficiency of 1 (E=1) corresponds to a perfect match of modeled discharge to the observed data. An efficiency of 0 (E=0) indicates that the model predictions are as accurate as the mean of the observed data, whereas an efficiency less than zero (-8 <E<0) occurs when the observed mean is a better predictor than the model. Essentially, the closer the model efficiency is to 1, the more accurate the model is.

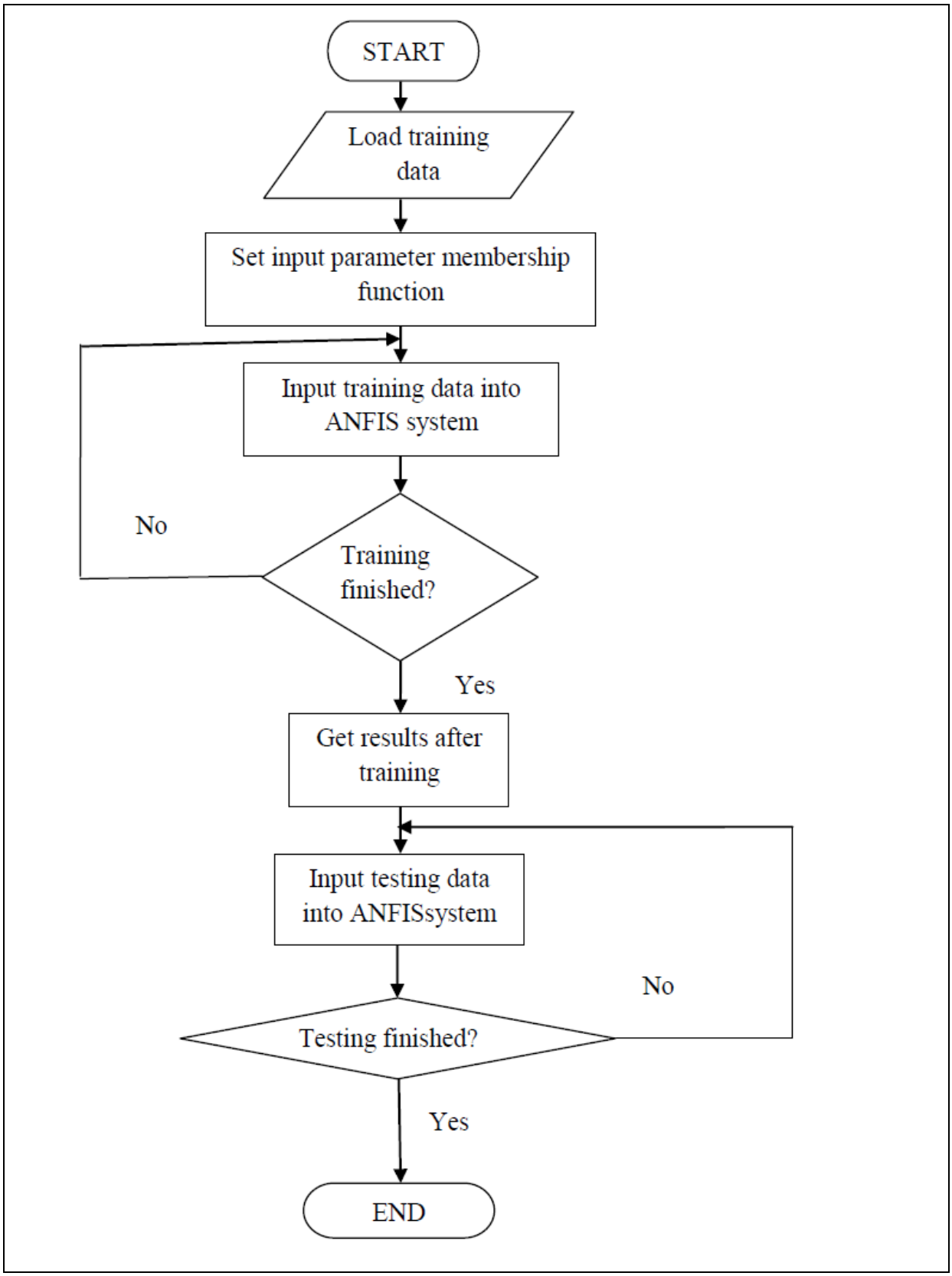


Figure 6.7: Flow chart showing steps of ANFIS model

Index of agreement (d): The index of agreement d was proposed by (Willmot, 1981) to overcome the insensitivity of E and r^2 to differences in the observed and predicted means and variances (Legates and McCabe, 1999). The index of agreement represents the ratio of the mean square error and the potential error (Willmot, 1984) and is defined as:

$$d = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}\left(\left|P_i - \overline{O}\right| + \left|O_i - \overline{O}\right|\right)^2} \tag{6.15}$$

The potential error in the denominator represents the largest value that the squared difference of each pair can attain. With the mean square error in the numerator d is also very sensitive to peak flows and insensitive for low flow conditions as it is E. The range of d is similar to that of r^2 and lies between 0 (no correlation) and 1 (perfect fit).

Correlation coefficient (R^2): The correlation coefficient (R^2), as given below:

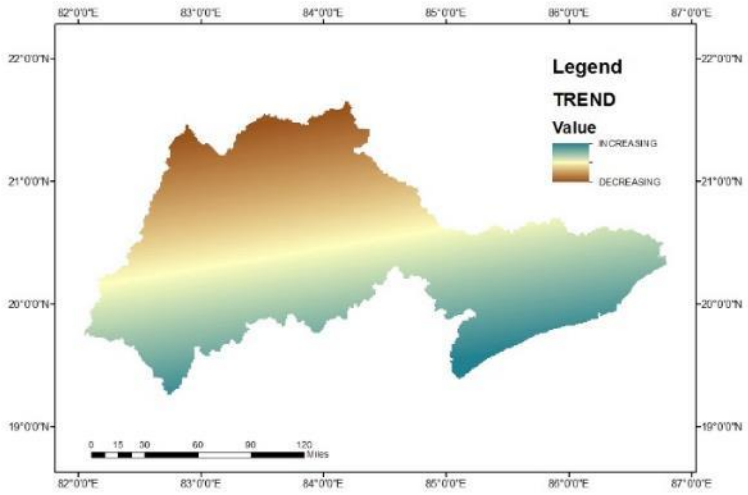
$$R^2 = \frac{\sum_{i=1}^{n}(Q_o - M_0)(Q_p - M_p)}{\sqrt{\sum_{i=1}^{n}(Q_o - M_o)^2 \sum_{i=1}^{n}(Q_p - M_p)^2}} \tag{6.16}$$

Where Q_0 and Q_P are the observed and estimated value, M_0 and M_P are the mean of the observed and estimated value respectively, and n is the total number of observations of the data set.

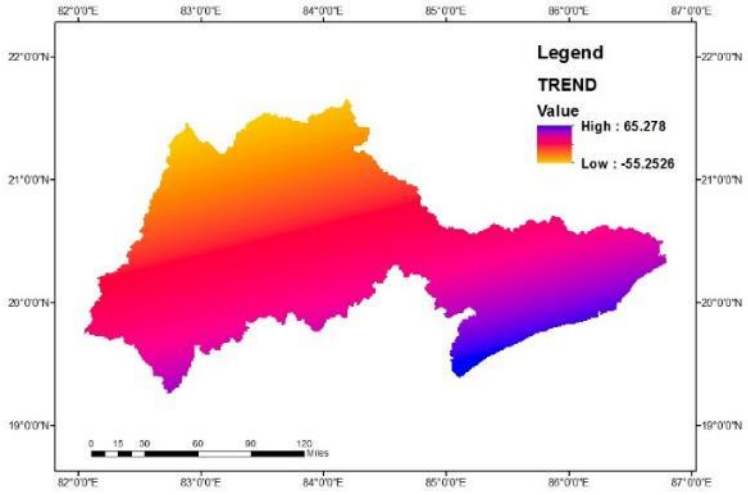
6.3 RESULTS AND DISCUSSION

6.3.1 Trend analysis

Mann–Kendall test and linear regression are used to detect the trend of different peak discharge of partial duration series having an average number of flood peaks per year (i.e λ=3) at Naraj, Hirakud, Kantamal, Salebhata for the period of 2000–2009 (Figure 6.8). Results from Mann-Kendall test shows the trend classification as increasing, decreasing and no trend. The Linear regression test shows trends as positive, negative and inconsistent slope. From both the methods, similar results are obtained, and the peak discharge is showing increasing trend at Naraj and decreasing trend at Hirakud of the lower Mahanadi basin.



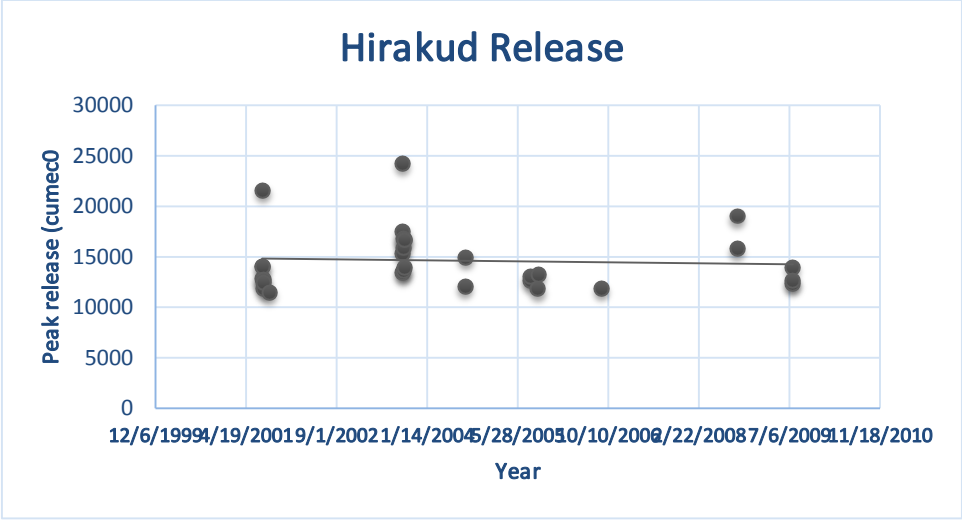
(a)



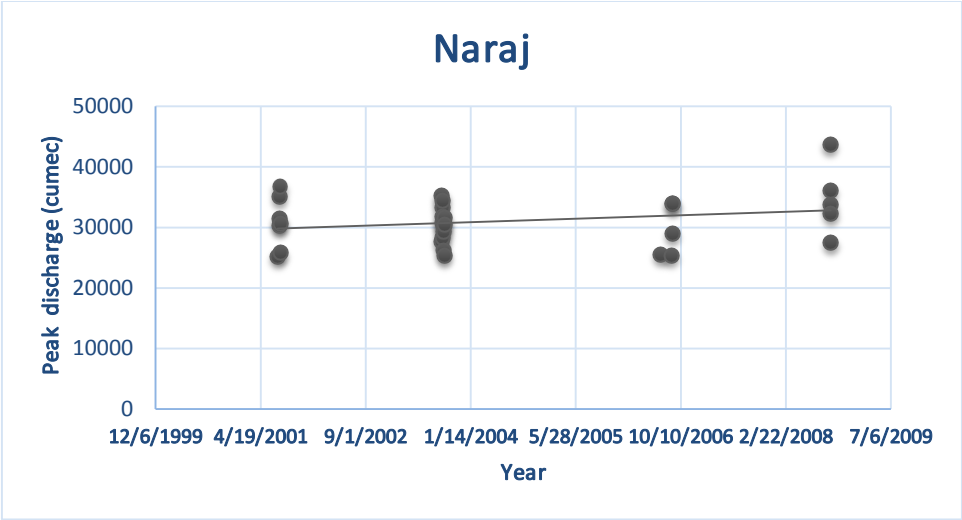
(b)

Figure 6.8: Trend analysis by (a) Mann-Kendall test and (b) linear regression for the peak discharge at Naraj, Hirakud, Kantamal and Salebhata in lower Mahanadi basin

Time series plots with the trend of peak release from Hirakud dam and peak discharge at Naraj for the period 2000-2009 are shown in Figure 6.9.



(a)

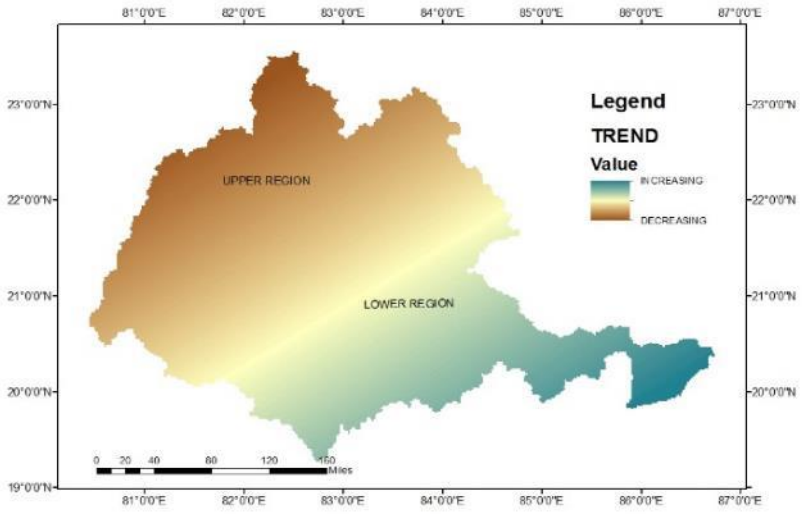


(b)

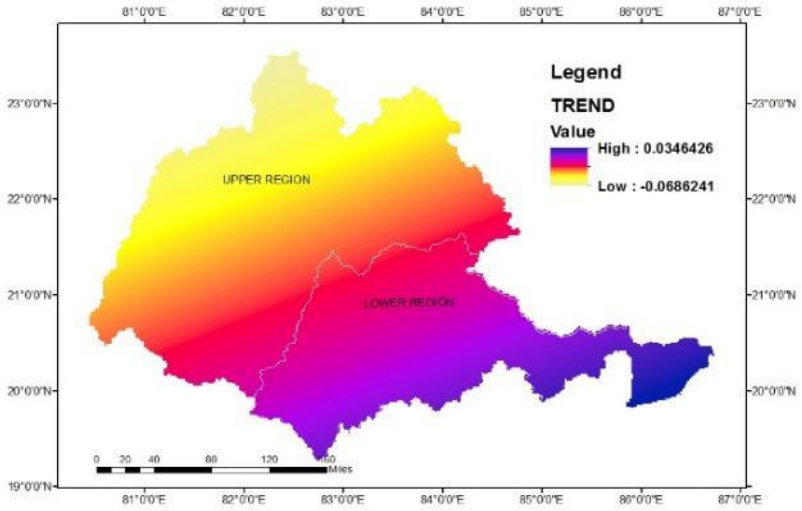
Figure 6.9: Time series plots with linear regression model (a) PDS discharge at Hirakud , (b) PDS discharge at Naraj

Figure 6.9 reveals that discharge obtained from partial duration series at Naraj in Mahanadi river basin exceeded a discharge of 28,400 m^3/s for twenty-one times. Amongst these, five flood events are very high and crossed 35,000 m^3/s during the years 2001, 2003 and 2008. Development of structural measures has been encouraged to control floods up to 35,000 $m^3/$ at Naraj. When it exceeds 35,000 m^3/s, non-structural mitigation measures are proposed as structural measure are not economical. It has been observed that the trend of peak discharge at Naraj is increasing whereas at Hirakud it shows the decreasing trend. This may be due to the increasing trend of both rainfall and peak flow contribution from the middle reach of the Mahanadi basin at Naraj. Hence, a region based (i.e upper and lower regions) analysis of

extreme rainfall is done. The results shows that the rainfall of lower region has an increasing trend whereas the upper region shows a decreasing trend in both the tests. Which confirm that the rainfall at lower region has also an impact on causing flood at the downstream site.



(a)

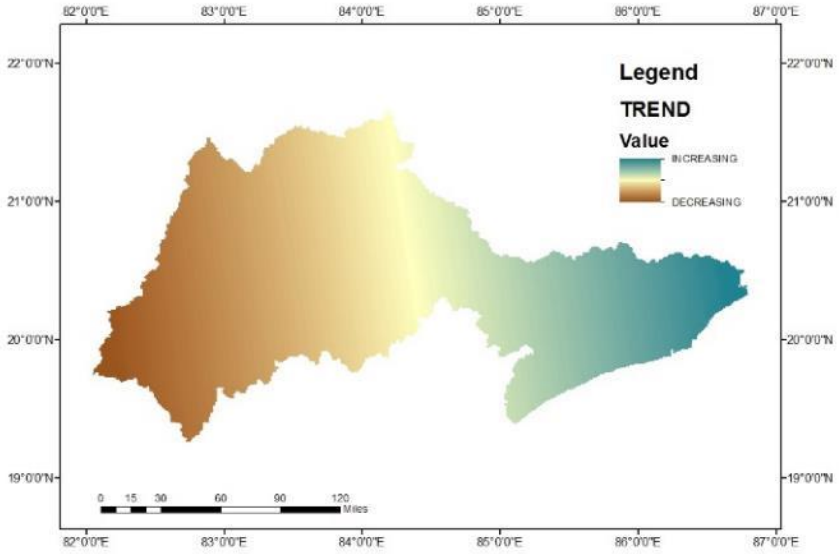


(b)

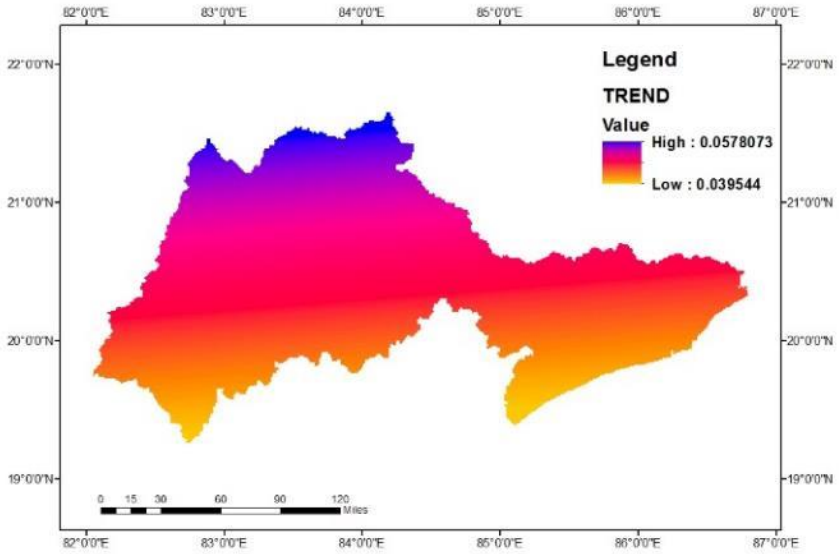
Figure 6.10: Trend analysis by (a) Mann-Kendall test and (b) linear regression for the rainfall extremes based on regions (upper and lower)

Further, the Mann–Kendall test and linear regression are also used to detect the trend of mean areal peak rainfall over the study area and are shown in Figure 6.11. It has been observed that

the peak rainfall is showing increasing trend in middle towards the Naraj station for all the cases.



(a)



(b)

Figure 6.11: Trend analysis by (a) Mann-Kendall test and (b) linear regression for mean areal rainfall in the lower region of Mahanadi basin

6.3.2 Random Number Generation

High-quality random numbers are generated from probability distributions function using Mersenne Twister algorithm in EasyFit software. Three hundred datasets were generated from 30 discharge datasets of partial duration series at different stations and used for further analysis.

6.3.3 Development of ANN and ANFIS Models

6.3.3.1 Model testing using ANN technique

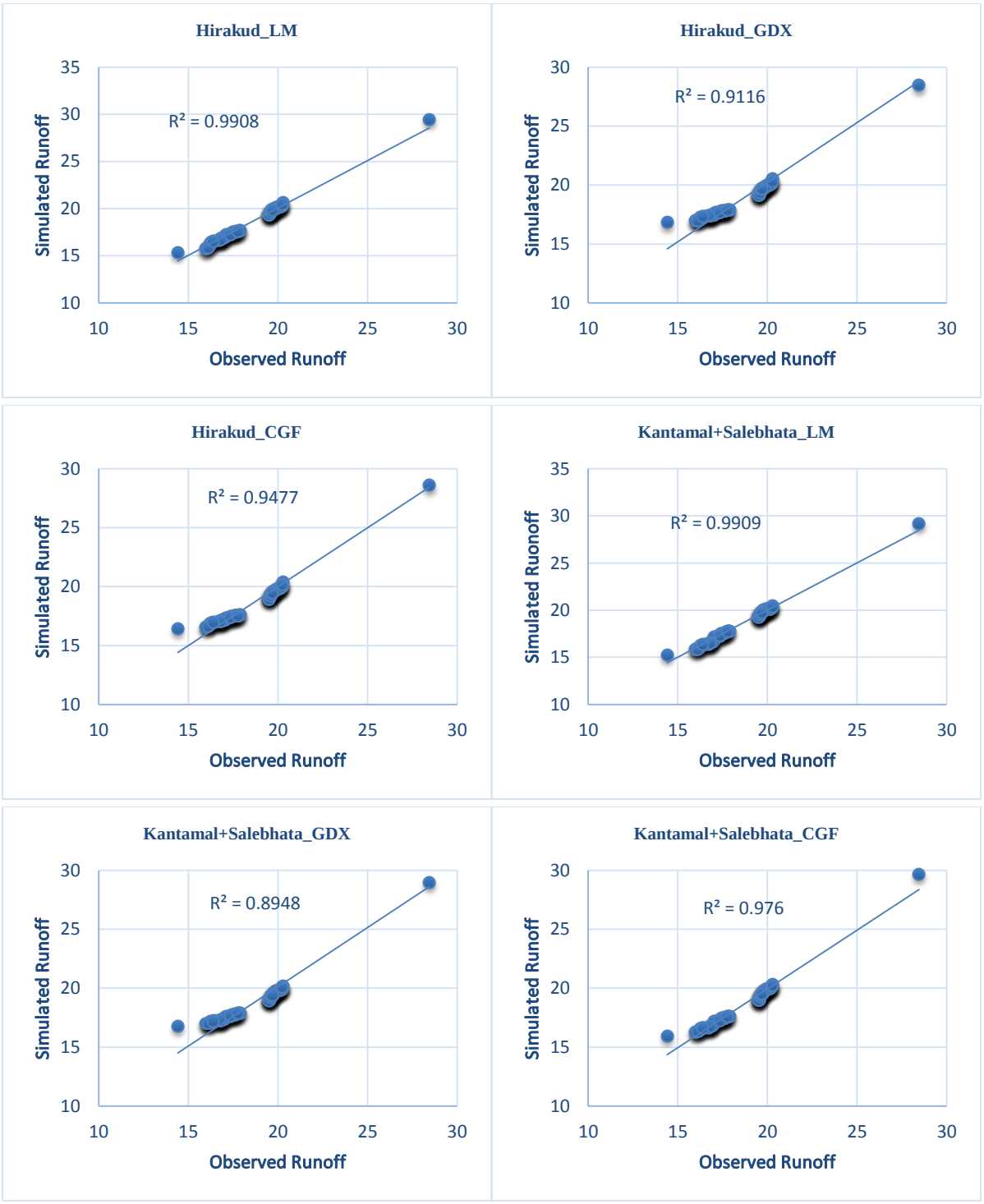
To run the ANN model, only one input parameter, namely the runoff from Hirakud Reservoir, was considered initially and subsequently input model parameters were increased to test the validity of ANN model with different input parameters. The combination of input variables used in ANN model is given in Table 6.1.

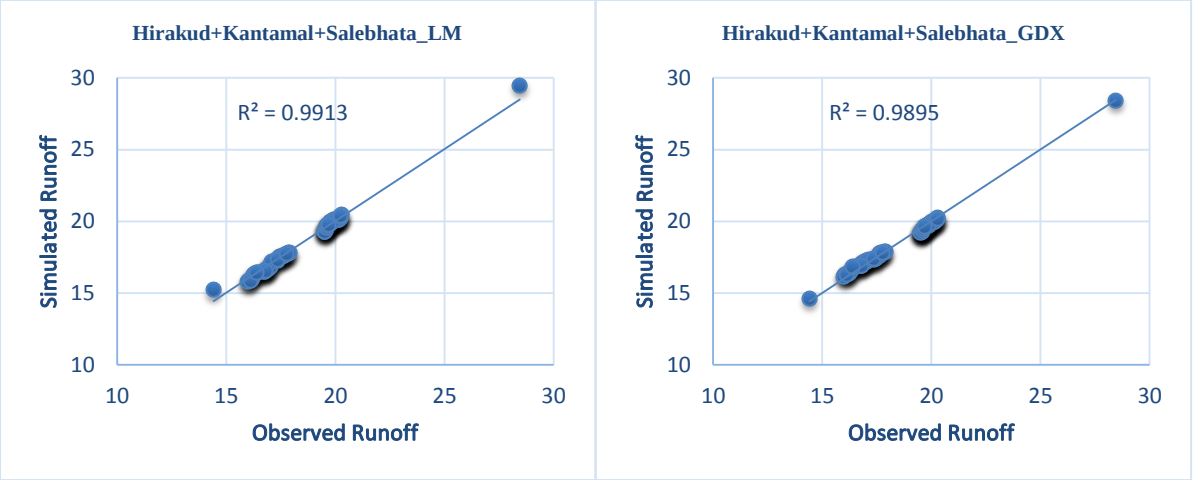
Table 6.1: Different combinations of input parameters in ANN modeling

Model	Input parameters	Output parameters
Model 1	Runoff at Hirakud	Runoff at Naraj
Model 2	Runoff at Kantamal+Salebhata	Runoff at Naraj
Model 3	Runoff at Hirakud+ Kantamal + Salebhata	Runoff at Naraj
Model 4	Rainfall at Lower Region	Runoff at Naraj
Model 5	Runoff at Hirakud+ Kantamal + Salebhata + Rainfall at Lower Region	Runoff at Naraj

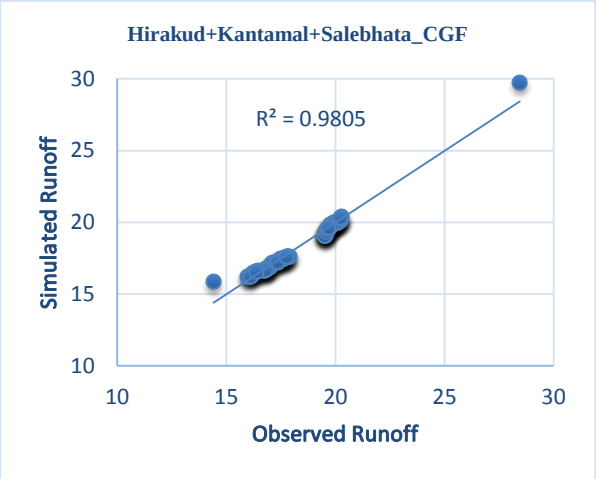
Training and testing of all the above models are carried out using a three-layered perceptron (Figure 6.12). The number of hidden neurons (HN) was varied till the best performance was acquired. Here hyperbolic tangent (tanh) and identity functions for the hidden and output neurons respectively have been chosen as activation function for all networks. The hyperbolic tangent (tanh) is a symmetric s-shaped (sigmoid) function, whose output lies in the range (-1, +1). With the identity function, the activation of the neurons is passed on directly as the output of the neurons, and the output lies in the range ($-\infty$, $+\infty$). Pre-processing of the datasets are made by normalizing them within the range of –1 to +1 and then putting into the ANN models for training and testing. To prevent the networks from overtraining and to enhance the generalization capability of networks, the training termination criteria used testing techniques to stop the training when the testing error began to increase. The number of maximum training epochs is set to 1,000, and the training is terminated when there is no further improvement in testing after 100 epochs. From 300 peak datasets, 210 datasets i.e., 70% are used to train and

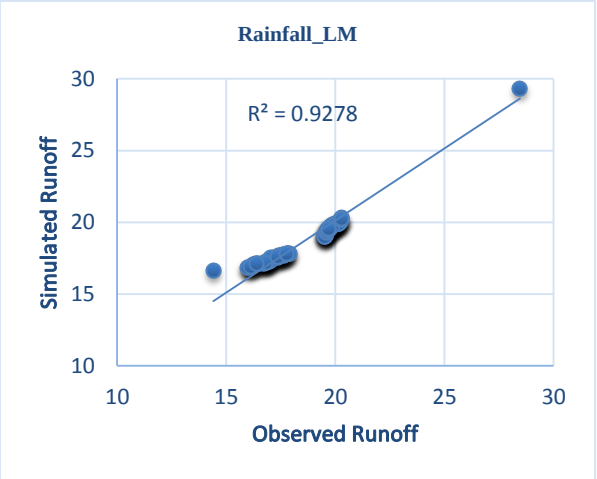
90 datasets i.e 30% are used to test the network. Three different back-propagation training algorithms, namely, the Levenberg-Marquardt (LM), the gradient descent algorithm with variable learning and momentum factor (GDX), and the conjugate descent algorithm (CGF), are applied to train the ANN models. The ANN simulation has been performed using Statistica software.

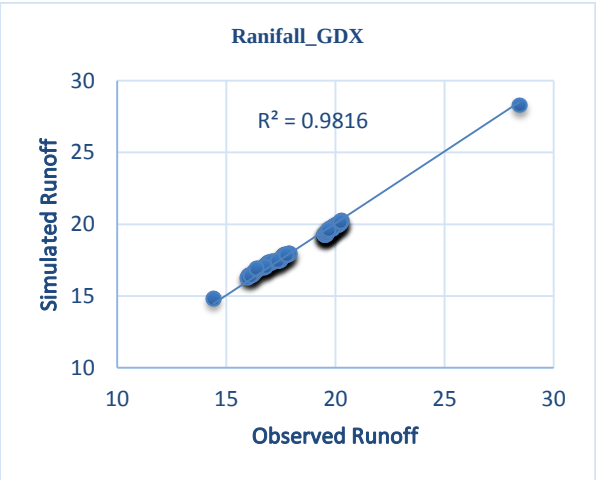


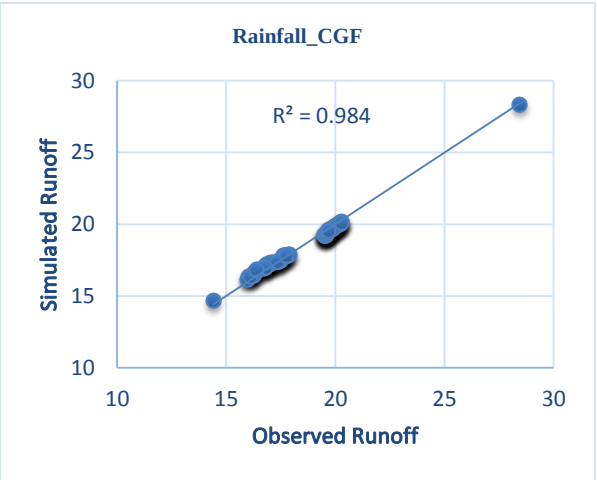

Hirakud+Kantamal+Salebhata_LM
R² = 0.9913
Simulated Runoff
Observed Runoff
10
15
20
25
30

Hirakud+Kantamal+Salebhata_GDX
R² = 0.9895
Simulated Runoff
Observed Runoff
10
15
20
25
30


Hirakud+Kantamal+Salebhata_CGF
R² = 0.9805
Simulated Runoff
Observed Runoff
10
15
20
25
30


Rainfall_LM
R² = 0.9278
Simulated Runoff
Observed Runoff
10
15
20
25
30


Ranifall_GDX
R² = 0.9816
Simulated Runoff
Observed Runoff
10
15
20
25
30


Rainfall_CGF
R² = 0.984
Simulated Runoff
Observed Runoff
10
15
20
25
30

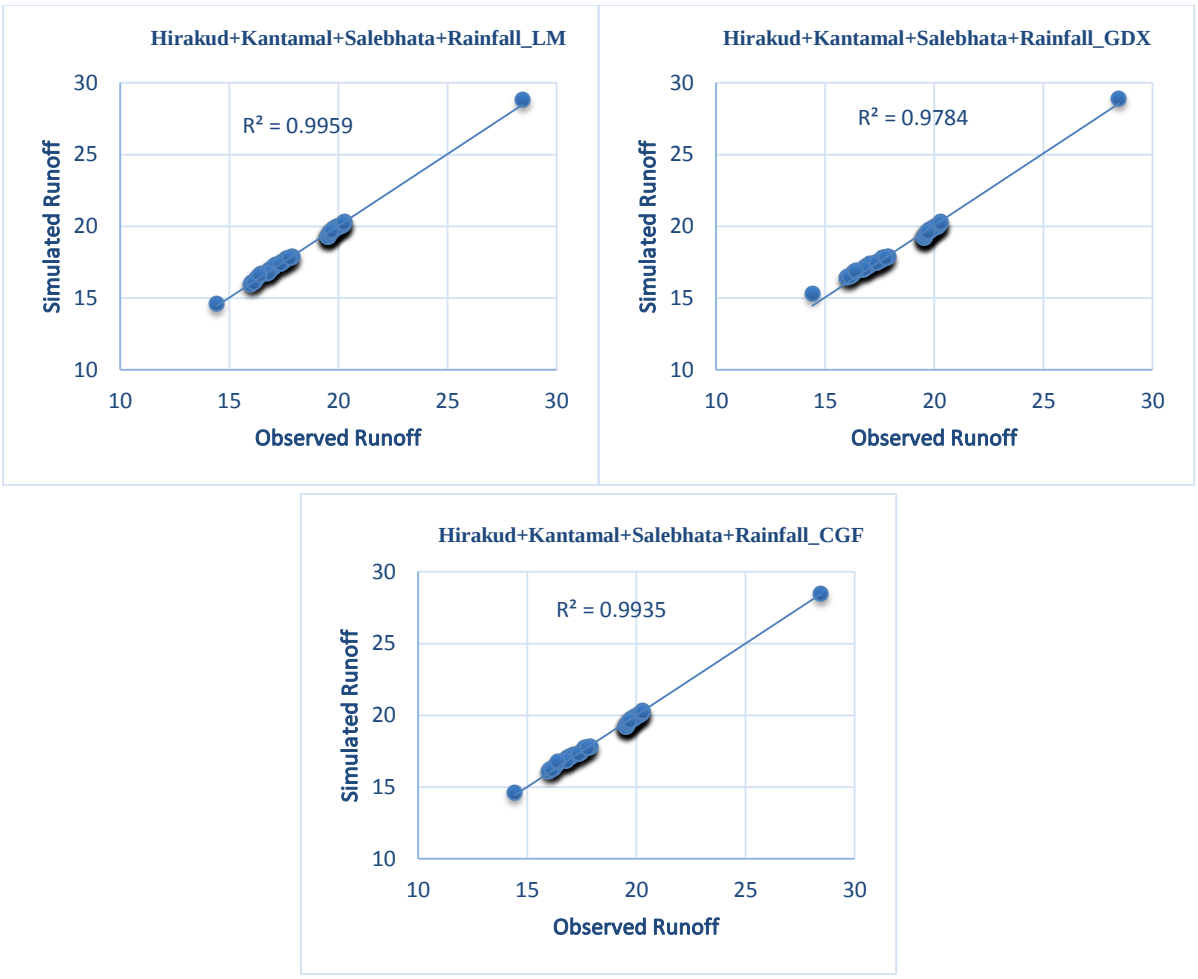


Figure 6.12: Comparison of observed and ANN-predicted runoff (trained using the LM, GDX and CGF algorithms respectively) during testing

Figure 6.12 show the observed and ANN computed flow during testing of all the input combinations using all three algorithm and found that model 5 having input combination of Runoff at Hirakud+ Kantamal + Salebhata + Rainfall at Lower Region has higher R^2 for LM algorithm in comparison to GDX and CGF algorithms. The ANN architecture with five hidden neurons was found to be proficient for simplifying input–output data sets.

The performance of all three algorithms are compared using root mean squared error (RMSE), Nash-Sutcliffe model efficiency (E), the index of agreement (d), and correlation coefficient (R^2). Figure 6.13 presents the performance of testing results obtained from all models trained by the three algorithms having a different number of hidden neurons (HN). It is appreciated that all models give satisfactory results presenting almost comparable values based on the performance criteria. However, Model 5 showed the best performance as evident from its highest E (0.996) and lowest RMSE (0.1201) values during testing. The Model 5 produced

minimum RMSE and higher E, IOA and R^2 using LM algorithm showing a relatively better results compared to GDX and CGF.

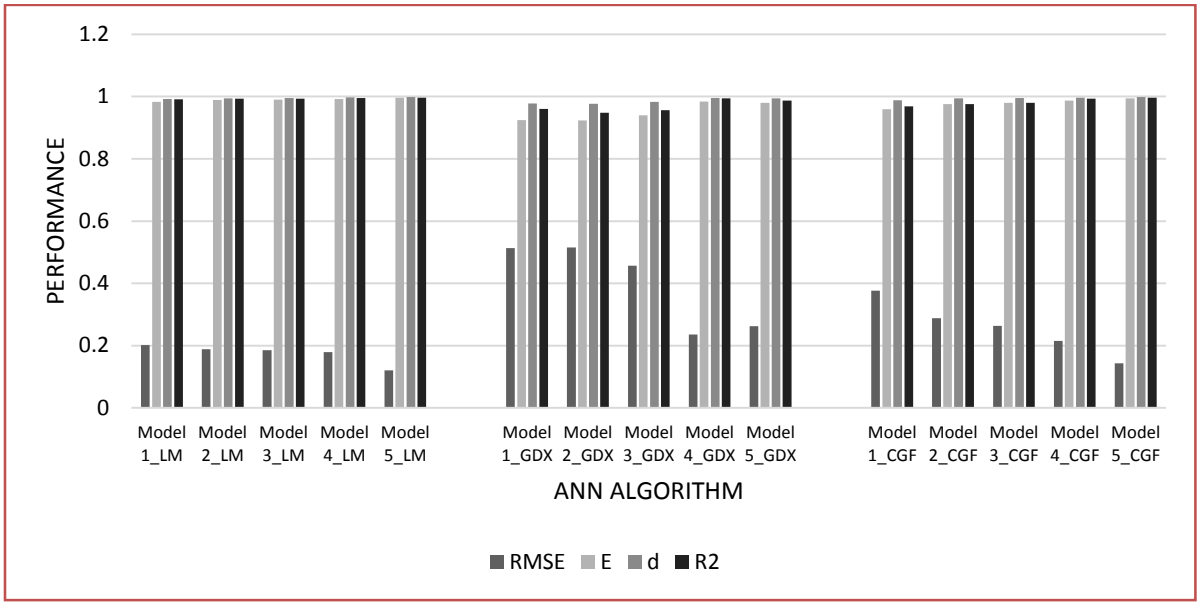


Figure 6.13: Performance evaluation of best Model-5 (using the LM, GDX and CGF algorithms respectively) during testing

6.4.3.2 Model testing using the ANFIS technique

A five-layered ANFIS model is created during training. Starting with two nodes, the number of nodes in the second layer is increased gradually during training of data. The error started decreasing with the increase of the nodes up to three. Hence, a number of nodes in the second layer is fixed to three and further analysis of ANFIS model has been carried out. The same combination of five input models are used in the ANFIS model too to predict runoff at Naraj. The data are normalized and scaled between 0 to1. The number of fuzzy membership functions for each input are considered as either 2 or 3 according to the type of model. The type of membership function used is of Gaussian type for inputs and linear type for output during generating fuzzy inference system. The number of fuzzy rules and the optimum number of parameters required for the best result are decided based on the number of inputs used, their type, as well as on the number of fuzzy membership functions employed in the model. The parameters of the membership functions are adjusted using the hybrid learning rule, the function ran steadily after 10 iterations. Figure 6.14 presents a comparison between the observed and ANFIS computed runoff obtained from all Model testing.

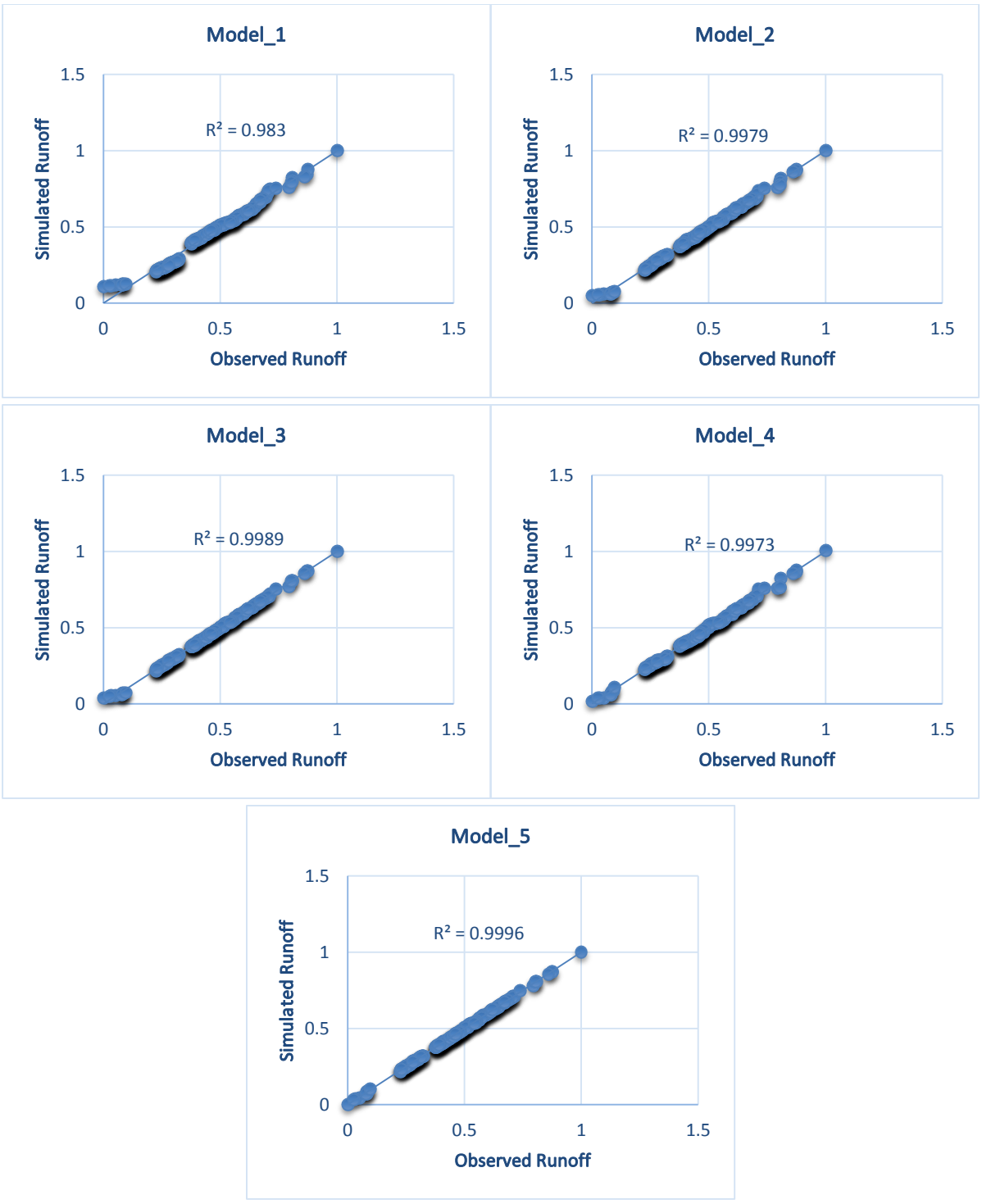


Figure 6.14: Comparison of observed and ANFIS-predicted runoff during testing

Figure 6.15 shows the performance indices obtained for all models. From the figure, it can be observed that Model 5 performed better than the other four models with E and RMSE values of 0.999 and 0.0168, respectively. The ANFIS generate quite good results, as shown by the

performance criteria for all models, compared to the ANN. The dominance of the ANFIS technique to the ANN method may be due to the fuzzy partitioning of the input space and for creating a rule-base to generate the output.

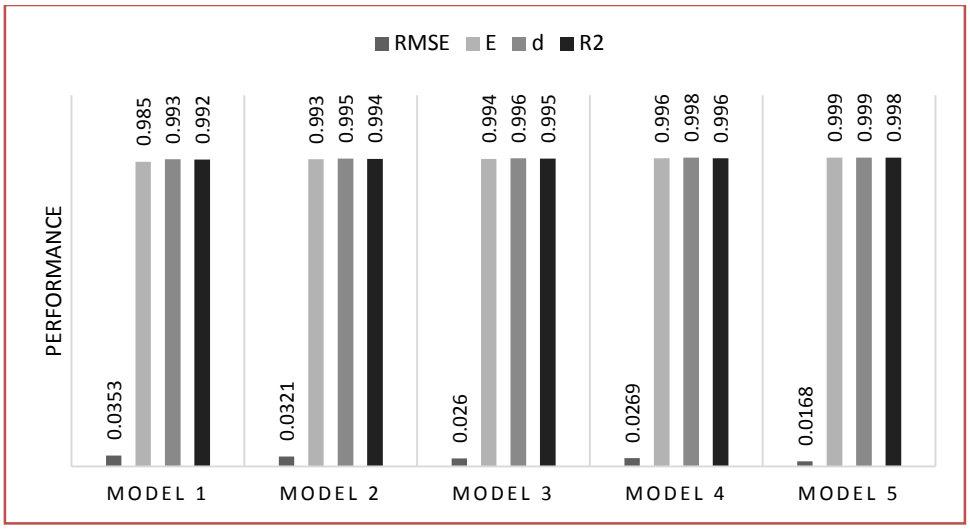
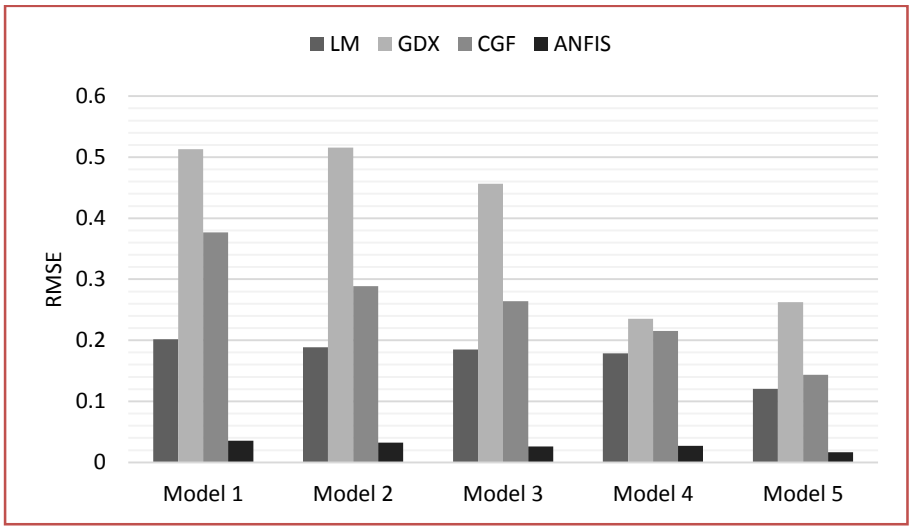


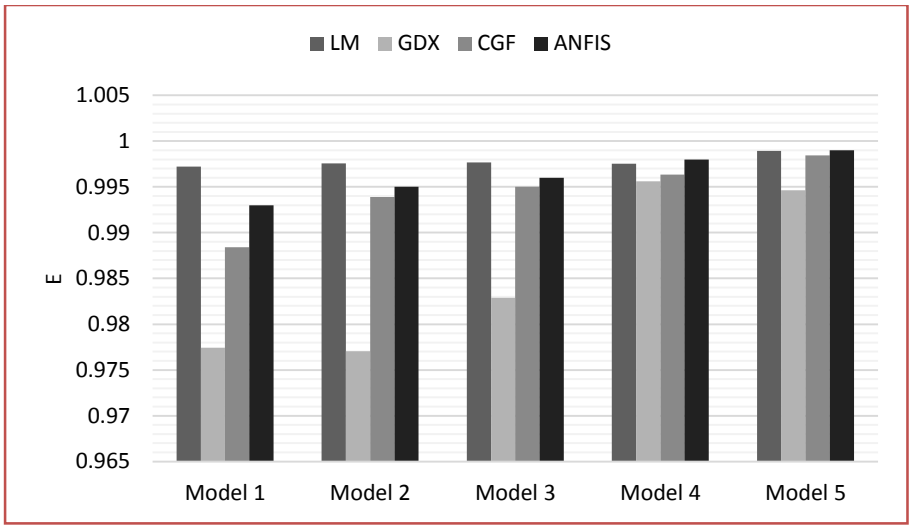
Figure 6.15: Performance evaluation of ALL Model during testing

6.4.3.3 Performance Evaluation of ANN and ANFIS

The performance of all three algorithms of ANN and ANFIS model was then compared mainly in terms of root mean squared error (RMSE) and The Nash–Sutcliffe efficiency (E) shown in Figure 6.16.



(a)



(b)

Figure 6.16: Comparison of RMSE and E values of Model 5 (ANN and ANFIS)

6.4 SUMMARY

This study presents a methodology for flood forecasting using partial duration series data having different sets of inputs and evaluate the effectiveness of the models at Naraj station. The trend analysis showed that the increasing trend of peak discharge at Naraj and the decreasing trend of peak releases at Hirakud recommend an increasing trend of both peak flow contribution and rainfall from the middle region of Mahanadi basin that may cause the flood problem situation at the delta region. The performances of all the three training algorithms of ANN: LM, GDX and CGF were found to be comparable. However, LM yielded the best result as revealed from the RMSE and Nash–Sutcliffe efficiency values. Five hidden neurons in one hidden layer were found to be the most appropriate in the ANN architecture yielding the best results. ANFIS performed slightly better than all ANN algorithms as revealed from the RMSE and Nash–Sutcliffe efficiency values.

CHAPTER 7
FLOOD INUNDATION USING 1-D HYDRODYNAMIC MODELING

7.1 INTRODUCTION

Flooding refers to the inundation of an area by unexpected rise of water which may occur due to dam failure or extreme rainfall or both. Most of the studies have applied hydraulic and hydrological models for simulating flood runoff in order to provide flood risk assessment information on the probability of flood occurrence, magnitude of the event, location and depth of the inundation for flood management. India mostly faces the flood problem during monsoon season, and it covers with clouds, therefore, the mapping of flood extent in optical remote sensing images is not clear, so, in this case, radar image can be preferable. The active Synthetic Aperture Radar (SAR) sensor allows acquisition of images independent of cloud cover, so in several studies, a combination of Geographic Information System (GIS), SAR imagery, high resolution digital elevation model (DEM) gives the flood inundation map with flood depth.

The flood problem in the delta region is generally serious due to heavy rains, dam releases due to heavy inflows and groundwater contributing to the river coming from the larger catchment area. In addition, due to the change of land use leads to the erosion of soils, which, in turn, are reducing the channel carrying capacity and reservoir capacity.

After commissioning of Hirakud dam during 1958 flood problems in Mahanadi system have been reduced significantly. However, it has been observed that it still continues to receive flood in the downstream area either as dam releases or as flow contribution from the middle regions below Hirakud. In Mahanadi system, mostly the rivers Kathjori, Devi, Kuakhai, Kushabhadra, Daya, Bhargabi, Birupa, Chitroptala, Paika drains most of its floodwater into the sea. Due to excess of water than carrying capacity of the rivers, major breaches occur on these rivers and the deltaic areas are always inundated during monsoon months. The most recent floods in the state occurred during September 2008, 2011 and 2013. While, the flood of 2008 in the Mahanadi basin was due to lower catchment contribution, the flood of 2011 was due to heavy rainfall in the upstream and the flood of 2013 was due to a severe cyclonic storm which was accompanied with torrential rains for 3 days.

The aim of this research is to determine the extent and depth of the flood at different return period using Hydrologic Engineering Center- River Analysis System (HEC-RAS) model and to check flooding risks in the delta of Mahanadi River by considering the main reach from Tikarapara station to Naraj station (Figure 7.1). Discharges with different return periods and

floodplain characteristics (situation of bed, river banks, etc.) are used in the HEC-RAS model to see how the discharge of different return periods at Naraj inundates the delta region of the Mahanadi river. Then, MODIS satellite data with its moderate- resolution optical sensor of 250-500 m is used to determine the inundated area.

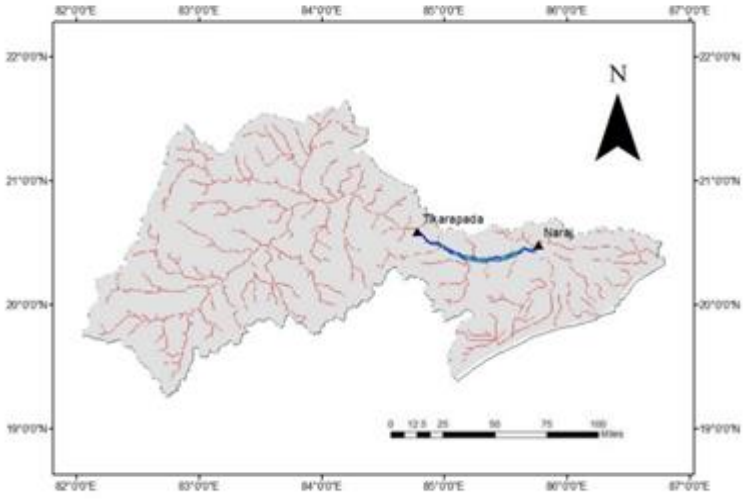


Figure 7.1: Index Map of Lower Mahanadi River basin, Odisha

The data used in this study are the time series of discharge and water level of different gauging stations, measured river cross-sections at different locations, topographical map, SRTM DEM and the daily surface reflectance MODIS data of 500-meter spatial resolution. The Channel geometry, boundary conditions, and channel resistance are required for conducting flow simulation through HEC-RAS.

7.2 METHODOLOGY

7.2.1 HEC-RAS and HEC-GeoRAS Model

HEC-GeoRAS is especially considered for processing the geospatial data. It is generally used to create a HEC-RAS import file containing geometric attribute data from an existing digital elevation model (DEM) and complimentary data sets. HEC-GeoRAS requires a Digital Terrain Model (DTM) of the river system in the form of a TIN or a GRID. In order to show the surfaces, both digital formats of Digital Elevation Model (DEM) and Triangular Irregular Network (TIN) were used. The geometric data developed in HEC-GeoRAS includes: stream center line, reaches (tributaries), cross-sectional lines, cross-sectional surface lines, cross-sectional bank stations, downstream reach lengths, main channel, right over bank, left over bank.

Once these layers are created successfully, the geometric data was exported to HEC-RAS as an import file for simulating the flood event. Land-use/land-cover data is required in shapefile

to generate Manning's n values. HEC-RAS model is used for computing one-dimensional steady flow which can simulate different water surface profiles. Information attained from channel geometry, and discharges values were used to create HEC-RAS channel flows. Through HEC-RAS, modeling of water surface profiles were developed from the corresponding discharge values. The HEC-RAS model calculates water surface profiles at all locations of interest at different return periods. After a successful simulation in HEC-RAS, we are considered the water surface profile at Naraj of different return periods and based on that we are generated the inundation map of Delta region.

7.2.2 HEC-RAS 1-D Hydrodynamic Modeling

The 1-D hydrodynamic model is based on the concept of flow in open channel theory. Open channel flow is defined as the fluid flow with a free surface exposed to the atmosphere, like the flow in natural streams, flow in drainage canals, and flow in storm sewers. The knowledge of the hydraulics of the open channel is required to the hydraulic engineer for the improvement of plans to successfully accomplish floodplain. In open channel flow, the continuity equation for steady flow relays flows to velocity and area. The equation states that flow must be conserved between adjacent cross-sections, the total energy per unit weight (energy head) has components of elevation head, pressure head, and velocity head and the momentum equation is expressed in the form of the Manning's equation. Water surface profiles are computed from one cross section to the next by solving the Energy equation with an iterative procedure called the standard step method. The Energy equation is written as follows:

$$H = y + z + \frac{\alpha V^2}{2g} \tag{7.1}$$

Where, H= energy head (m); Z= base channel elevation (m); α= velocity weighting coefficient

Based on these parameters, the water surface elevation is the sum of y and z. The change in energy head between adjacent cross sections is equal to the head loss:

$$H_1 = H_2 + h_L \tag{7.2}$$

Where, H_1 = energy head at cross section 1 (m); H_2 = energy head at cross section 2 (m); h_L = energy head loss (m)

The head loss between the two cross sections is the sum of friction head loss and flow contraction/expansion head loss. Friction losses result from shear stress between the water and channel bottom:

$$h_f = L\overline{S_f} \tag{7.3}$$

Where, h_f = friction head loss (m); L = distance between adjacent cross sections (m);
Contraction/expansion head losses can occur as a result of the formation of eddies wherever there is a contraction or expansion of the channel:

$$h_o = C\left|\frac{\alpha_2 V_2^2}{2g} - \frac{\alpha_1 V_1^2}{2g}\right| \quad (7.4)$$

Where, h_o = contraction or expansion head loss (m); C = contraction or expansion coefficient
Consider the Figure 7.2 the flow from section 2 to section 1 must be conserved.

$$Q = V_1 A_1 = V_2 A_2 \quad (7.5)$$

Where, Q = flow rate/discharge (m^3/s); V_n = average velocity at cross-section n (m/s); A_n = area at cross-section n (m^2)
And conveyance (K) is calculated from the following form of Manning's equation:

$$Q = KS_f^{1/2} \quad (7.6)$$

and

$$K = \frac{1}{n} AR^{\frac{2}{3}} \quad (7.7)$$

Where, K = conveyance ($m^{8/3}$); S_f = average friction slope; n = Manning roughness coefficient
R = hydraulic radius (m)
The Manning's roughness coefficient (n) is a parameter that measures the effect of the roughness of the channel on the flow of water through it. The values, which vary based on channel terrain, are published in most hydraulic engineering books. The hydraulic radius (R) is calculated by dividing the cross-sectional area by the wetted perimeter. The value of channel slope (S) is derived from DEM.

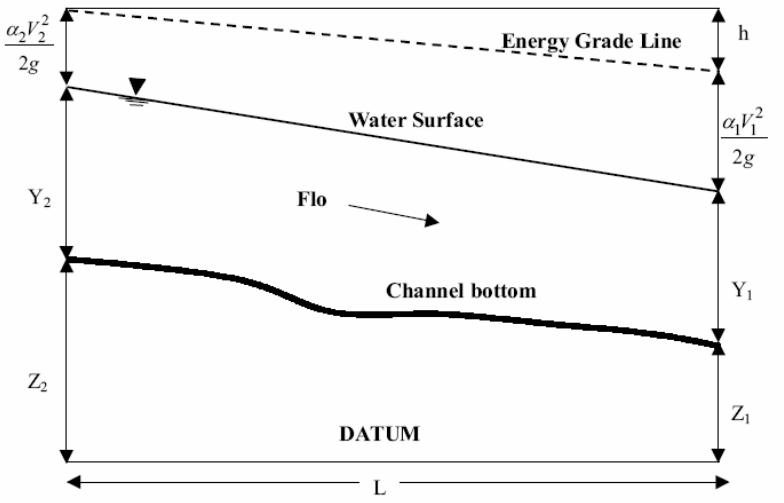


Figure 7.2: Representation of Energy equation (Source: HEC-RAS User Manual, 2002)

7.2.3 Flood Inundation Mapping

The 500-meter spatial resolution of daily Surface Reflectance MODIS data makes a natural tool to monitor soil moisture conditions and has been used as input to calculate the Normalised Difference Water Index (NDWI) model. The NDWI is generally used to identify water surface (Cretaux et al., 2011). Before processing the analysis, a data quality control process was applied to screen the cloud and fill value pixels obtained from the associated MODIS quality assurance (QA) data product. The main reason of using NDWI is that short-wave infrared (SWIR) radiation is highly sensitive to moisture content in the soil and the vegetation canopy. A number of studies have been conducted in use of the spectroscopic characterization of SWIR to detect water content of an area (Gao, 1996). Cloud cover is always an issue when trying to map flood events using optical remote sensing, especially during the rising stage of a flood event. To help reduce the effects of cloud interference, the use of 8-day (MOD09A1) or 16-day (MCD43A4) MODIS composites (provided by NASA) has proven to be useful in mapping the temporal dynamics of water (Ordoyne and Friedl, 2008; Chen et al., 2011; Weiss and Crabtree 2011; Chen et al., 2012; Huang et al., 2012a; Huang et al., 2012b; Chen et al., 2013).

$$Normalised Difference\ Water Index\left(NDWI\right) = \frac{Band_{RED} - Band_{SWIR}}{Band_{RED} + Band_{SWIR}} \tag{7.8}$$

where $Band_{RED}$ is the reflectance of Red [621–670 nm], MODIS Band1 and $Band_{SWIR}$ is reflectance of short wave infrared [1628–1652 nm], MODIS Band 6 of the solar spectrum.

After identifying water-related pixel using NDWI, it is essential to classify whether it is a Flood pixel or it is a water body. For that a pre-flood image of MODIS data of different return periods has been used. All Image Processing work, Geoprocessing and output maps are prepared on ARC GIS 10.2 platform. The algorithm for processing of NDWI has been done in Model Builder on ARC GIS 10.2.

7.3 RESULTS AND DISCUSSION

The Digital Elevation Model (DEM) for the River Mahanadi derived from SRTM data having 90 m spatial resolution has been used as an input for terrain processing (Figure 7.3). ArcGIS software is used to generate Triangulated Irregular Network (TIN) from DEM as shown in Figure 7.4, which is the required input data for simulating HEC-RAS model. The geometric data developed in HEC-GeoRAS includes; stream center line, reaches, cross sectional lines,

cross sectional surface lines, cross sectional bank stations, downstream reach lengths, main channel, right overbank, left overbank (Figure 7.5).

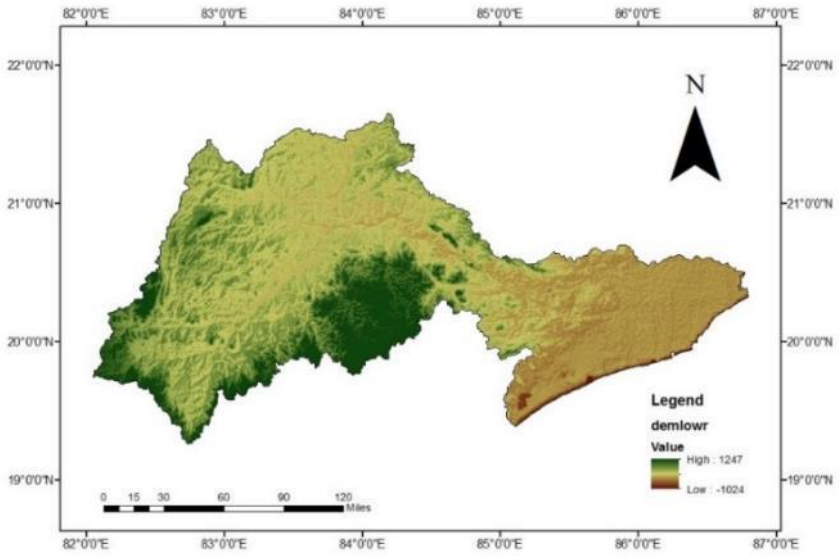


Figure 7.3: Digital Elevation Map of Lower Mahanadi River basin, Odisha

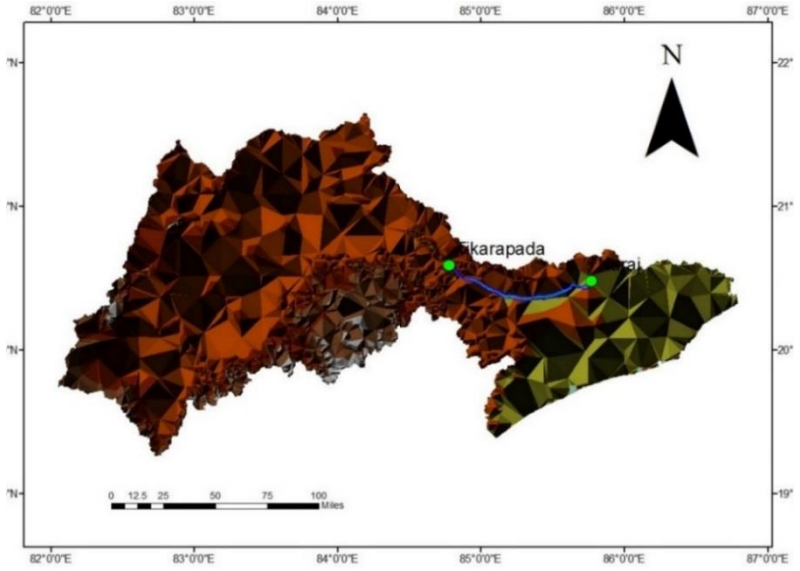


Figure 7.4: Digital Terrain Map of Lower Mahanadi River basin, Odisha

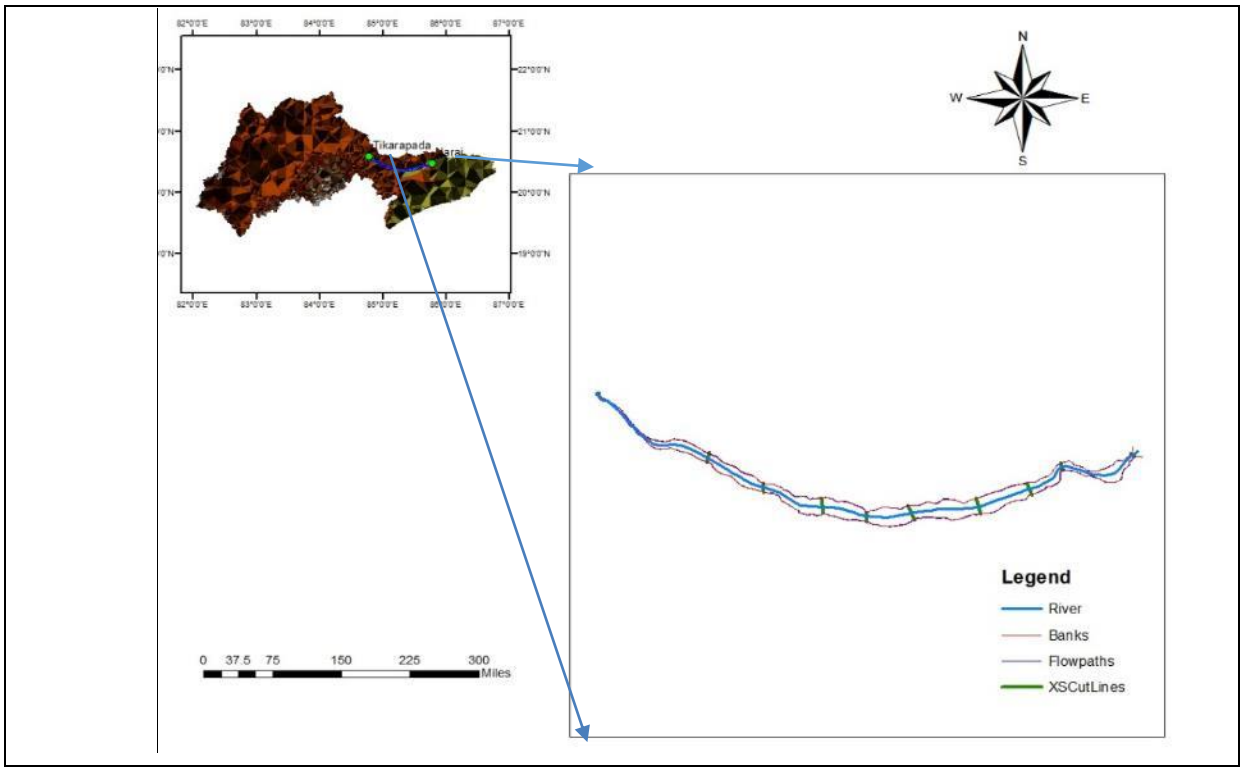


Figure 7.5: Geometry Map created in HEC-GeoRAS

The inputs data files, i.e., river network, cross-section and storage area are created in HEC-GeoRAS and exported in HEC-RAS for hydrodynamic simulation. The 1-D Hydrodynamic modeling is done in HEC-RAS to generate water surface profiles at different cross-section for different return period like 2, 5,10, 20, 50 and 100 years. Water surface elevation data is used as the upstream boundary condition and the normal depth is used as the downstream boundary condition. This boundary condition requires the input of the Energy Grade Line (EGL) slope at the downstream boundary. The water surface elevation profiles and rating curve at Naraj gauge site for different return periods are shown in Figure 7.6 and 7.7. It is relevant to mention here that the warning level at Naraj is 25.41m and danger level is 26.41m. From the Figure it is shown that for all return periods, the water level crossed the danger level and that excess amount of discharge creates the flood problem in the downstream sections.

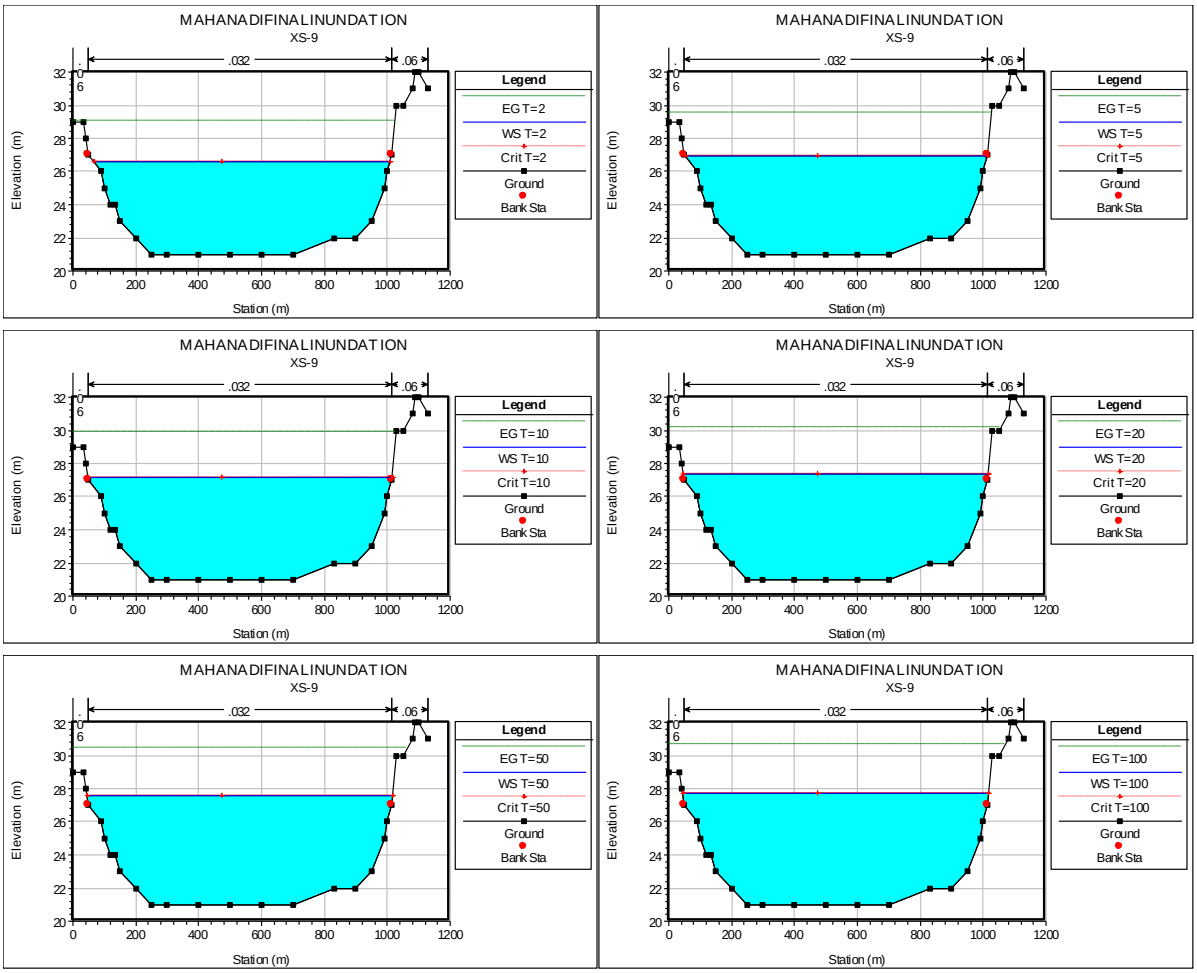


Figure 7.6: Water level at Naraj Station for different Return periods (2, 5, 10, 20, 50 and 100 years)

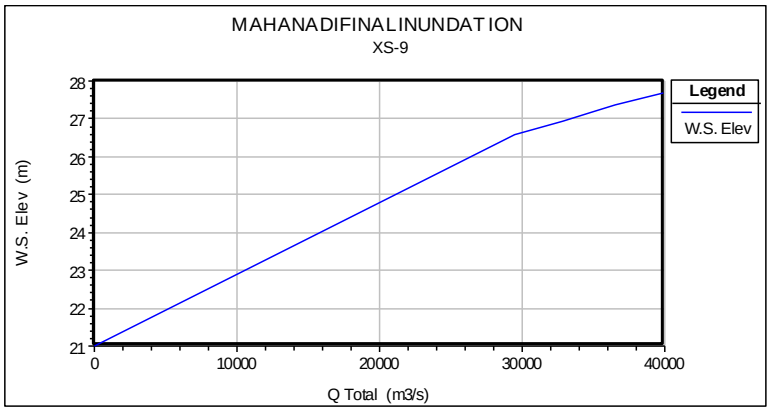


Figure 7.7: Rating Curve

The estimated peak discharge values for the return period 2, 5, 10, 20, 50 and 100 are found to be 29463.91m^3/s, 32867.42 m^3/s, 34853.82 m^3/s, 36581.27 m^3/s, 38584.68m^3/s and 39931m^3/s.

From the observed data, it is identified that (i) the 2 and 10 year return period flood occurred during the month of August in the year 2003, (ii) the 5, 50 and 100 years return period flood occurred during the month of October in the year 2008, and (iii) 20 year return period flood occurred during the month of July in the year 2001. Hence, Surface Reflectance MODIS data for the above period are downloaded. The red and short-wave infrared (SWIR) band data are extracted for calculation of NDWI by using Model Builder in ARC GIS (Figure 7.8) to identify the surface water content. After identifying water-related pixel, reclassification is done to know whether it is a flood pixel or a water body (i.e., pond). For that, pre- flood images of the above year were used to outline all long term water bodies (Figure 7.9).

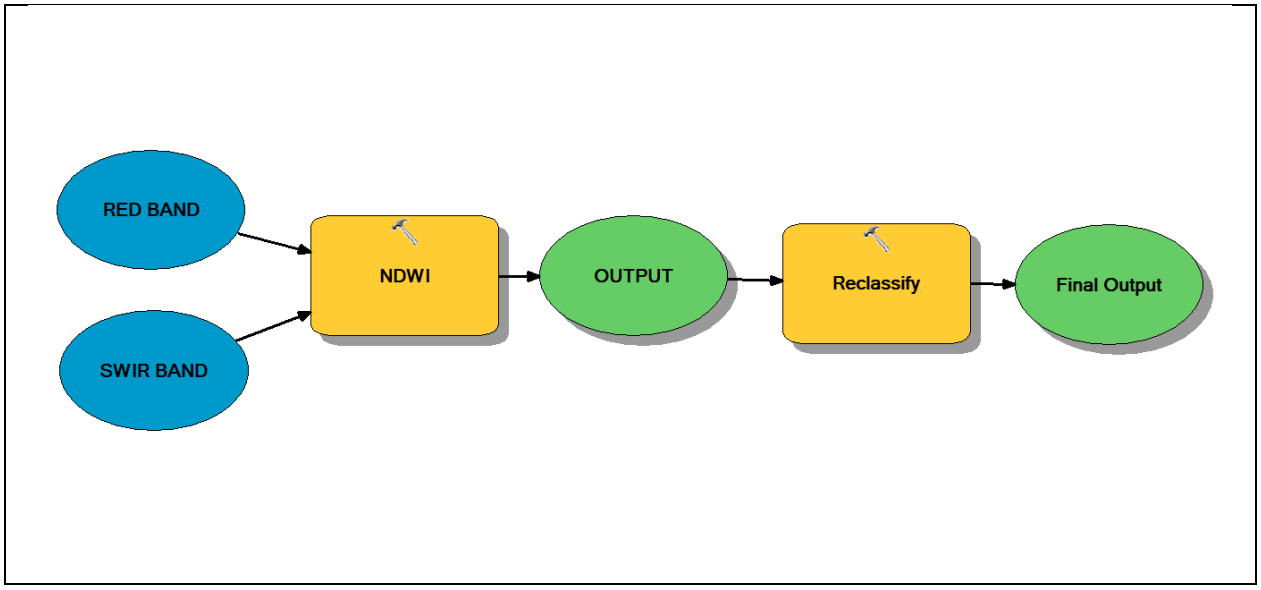


Figure 7.8: NDWI Model

Using this proposed methodology, changes of spatial extent with time are analysed and flood inundation maps are developed (Figure 7.10, 7.11 and 7.12) for different return periods. It has been observed that the inundation areas for 20, 50 and 100 year return periods exceed the discharge value of 35000 m^3/s at Naraj, and largely affects the agricultural lands and communities located in the delta region of Mahanadi River basin. There are large numbers of built-up areas and urban settlements which are at a risk of inundation. It also shows how the high risk area is flooded by the event.

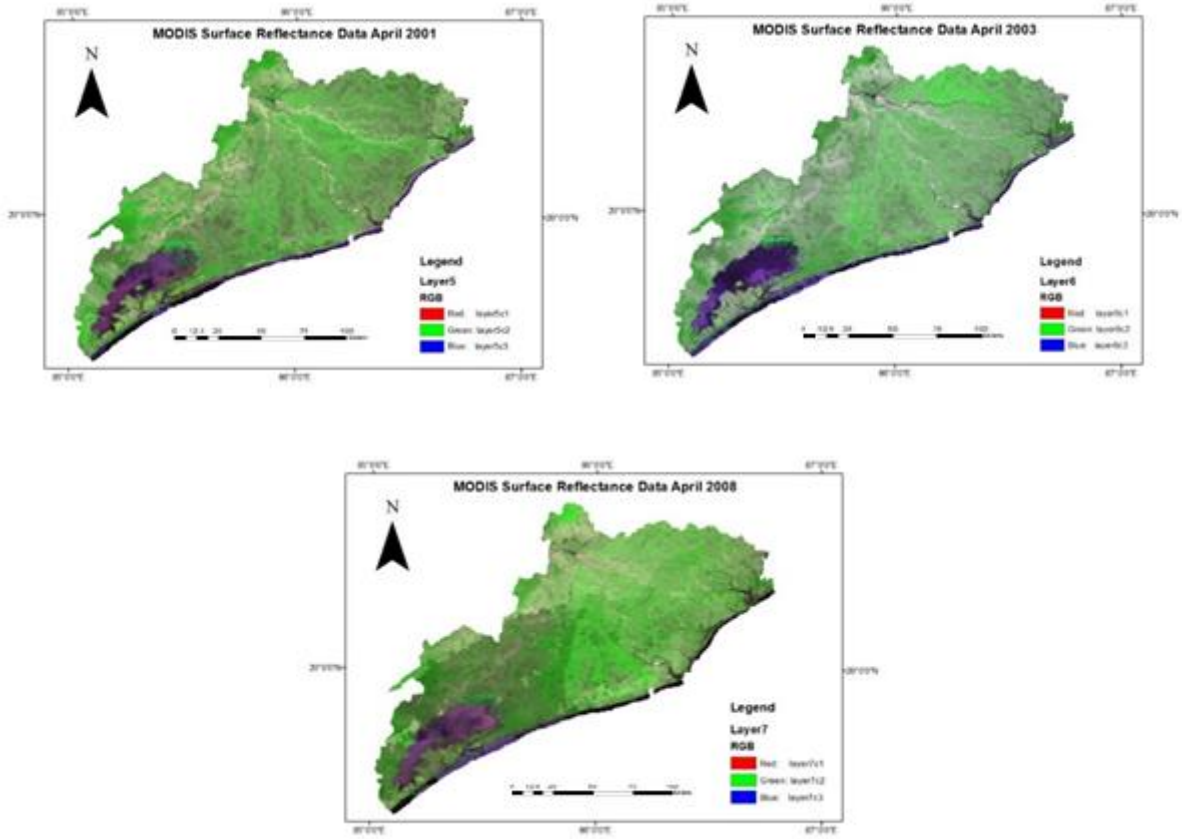


Figure 7.9: Observed MODIS Surface Reflectance Data (April 2001, 2003 and 2008)

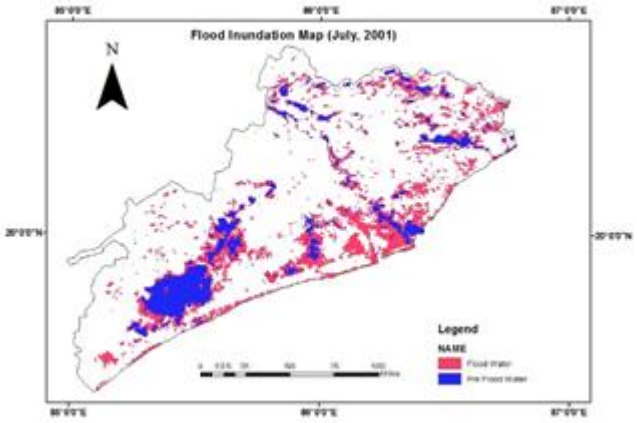


Figure 7.10: Flood inundation maps using MODIS Data (20 Year Return Period)

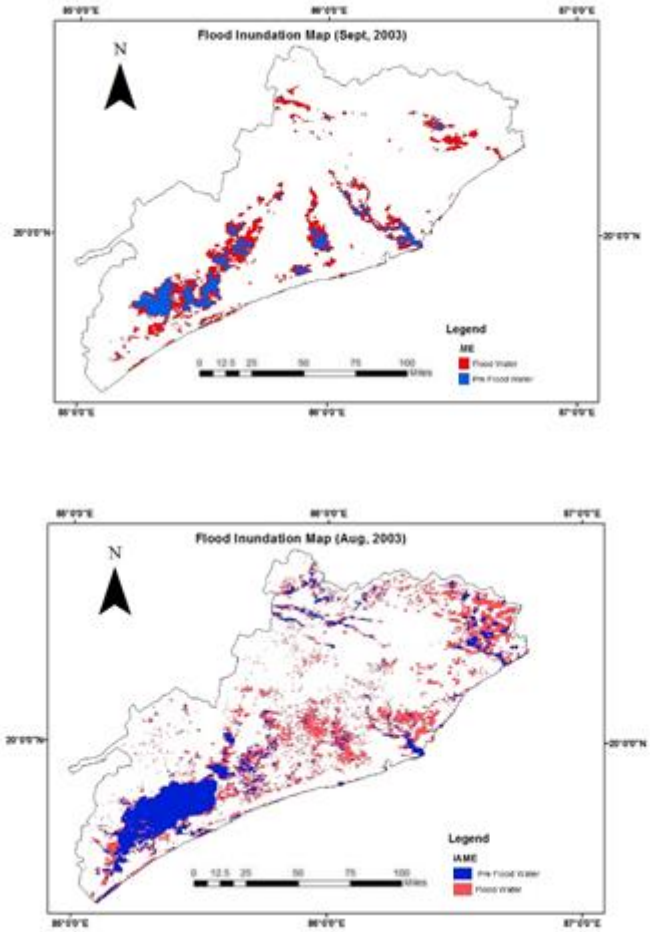


Figure 7.11: Flood inundation maps using MODIS Data (2 and 10 Year Return Periods)

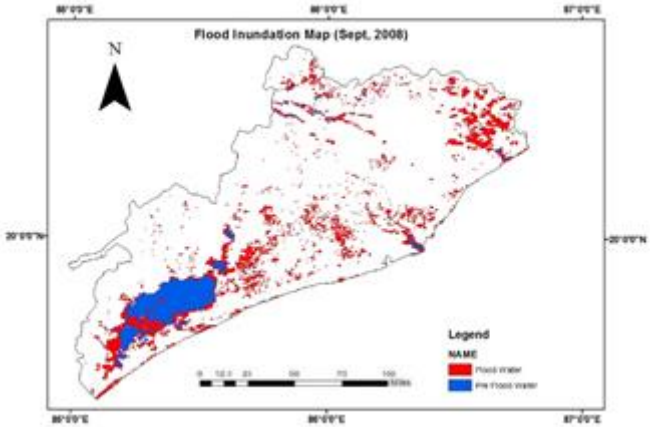


Figure 7.12: Flood inundation maps using MODIS Data (5, 50, 100 Year Return Periods)

7.4 SUMMARY

This research study has demonstrated the application of HEC-GeoRAS, HEC-RAS in GIS environment to identify and recognize flood inundation areas for flood risk management. HEC-RAS model has been calibrated to delineate flood inundation areas for different return periods for the Mahanadi river basin. Hydraulic modeling using GIS technique has proved quite useful in simulating flood water depth and inundation areas for various return periods of Mahanadi River basin. MODIS satellite images have been used to develop flood inundation maps. Such inundation maps are found to be useful for integrating water resources management and the maintenance of ecosystems of wetlands.

CHAPTER 8
CONCLUSIONS & RECOMMENDATIONS FOR FUTURE WORK

8.1 CONCLUSIONS

In the present study efforts have been made to (a) estimate threshold values for partial duration series of peak flows in Mahanadi river basin using different approaches and derive a flood frequency model for different return periods at each sampling station, (b) develop regional flood formulae using partial duration peak flows for the entire Mahanadi basin, (c) develop flood forecasting models using soft computing techniques like ANN and ANFIS for the lower region of Mahanadi basin and, (d) use HEC-GeoRAS and HEC RAS models to assess the flood inundation due to peaks flows of different return periods. The conclusions drawn from this study are summarized in following sections:

8.1.1 Selection of Threshold in Partial Duration Series Modeling

The application of the PDS for flood frequency analysis requires the selection of threshold level (t) or average peaks per year (λ). In fact, no unique specific value can be selected for the entire basin due to different physiographic characteristics at various locations within the basin. In the present work, an attempt is made to evolve the operational guidelines for selection of threshold values for obtaining partial duration series (PDS) for 22 stations located in the Mahanadi river basin, India. The results indicates that the threshold level having average peaks per year (λ) ranges in between 2 to 3 in all the cases. The PDS obtained for each station satisfies all the tests, assumptions and hypothesis.

The suitability of the GP distribution is also inspected using the L-moment ratio diagram and stability test of GP parameters coupled with quantile estimation. For small λ values (higher threshold), a variety of distributions are found be fitted best, for the λ equal to or greater than two, only the GP distribution provides good results. The stability of the GP parameters with different threshold coupled with the quantile estimation is also checked.

The results confirmed that the Generalized Pareto (GP) best described the PDS in the study area and also, show that the best PDS/GP performance is found in almost all the value of λ (2, 2.5 and 3). Therefore, in the application of PDS in the Mahanadi river basin, the threshold value at different stations with the λ values ranging in between 2 to 3 are preferred. From QQ plots it has been observed that at lower λ values (higher threshold), the higher quantiles are under-fitted and at higher λ values (lower threshold), the models tend to over-fit the higher quantiles.

The quantiles obtained from higher return periods are much influenced by the shape (ξ) and scale (σ) parameters of GP distribution model but for smaller return periods it is not.

8.1.2 Regional Flood Frequency Analysis of Partial Duration Series

Regionalization is the best and viable way of improving flood quantile estimation. In the regional flood frequency analysis, selection of basin characteristics, morphology, land use and hydrology have significant role in finding the homogeneous regions. The Mahanadi river basin is divided into two statistically homogeneous regions by using nine input variables obtained by factor analysis and clustering techniques. There is no major influence on homogeneity and cluster formation by reducing the dimensionality of variables from eleven to nine. The homogeneity tests (discordance measure test, Cv and LCv based homogeneity tests and statistical comparison), for all stations have satisfied the criteria except two stations in both the regions.

The K-Mean (KM) clustering method provides better results in comparison to Hierarchical Clustering (HC) method. Here, again the Generalized Pareto distribution (GP) best described the PDS in the study area. To test the homogeneity and to identify the best-fit frequency distribution, regional L-moment algorithm is used. A unique regional flood frequency curve is developed which estimated the flood quantiles of ungauged catchments and an index flood is also specified by using the multiple linear regression approach.

8.1.3 Application of Soft Computing Techniques for River Flow Prediction in Lower Mahanadi River Basin Using Partial Duration Series

The trend analysis showed increasing trend of peak discharge at Naraj and the decreasing trend of peak releases at Hirakud. To demonstrate the rainfall, and corresponding peak flows obtained using PDS, ANN and ANFIS models are used successfully. It has been observed that the floods and inundation due to these peak rainfall typically depend on various parameters including time of concentration, basin slope, river morphological characteristics, rainfall, soil moisture, groundwater, land use, and river discharge during monsoon. For this reason, a different combination of parameters have been used to obtain the best model.

The runoff at Naraj station in the Mahanadi River, India, has been predicted with considerable accuracy by taking the Hirakud runoff, runoff at Kantamal and Salebhata, and the mean rainfall of middle region (Model 5) as inputs to the ANN and ANFIS techniques. ANFIS performed better than ANN algorithm as revealed from the RMSE and Nash–Sutcliffe efficiency values.

8.1.4 Flood Inundation Using 1-D Hydrodynamic Modeling

The research study has demonstrated the application of HEC-GeoRAS, HEC-RAS in GIS environment to recognize flood inundation areas for disaster risk management. The GIS aided in data informing, visualization of changed scenarios and risk mapping. HEC-RAS model is calibrated to measure flood inundation at different return periods for the Mahanadi river basin.

Due to increase in flooding frequency, population residing near the river banks, agricultural land and other valuable infrastructure like roads and bridges are found at high risk of flood inundation. Hydraulic modeling using GIS technique proved useful in simulating flood water depth and inundation areas for various return periods of Mahanadi River basin. Flood inundation maps are developed from Normalized Difference Water Index (NDWI) derived from MODIS surface reflectance data and has been compared with SRTM data to understand the flood area. The products derived from MODIS 500m imagery shows the ability to study flood dynamics considering that MODIS products have a great advantage in the high-frequency flood observation.

8.2 RECOMMENDATION FOR FUTURE WORK

Based on this study, following recommendations is made for further work in the area. The recommended directions in which further work can be undertaken are listed below:

i) One of the limitations of the study is the data deficiency, hence more stations with more years of data give accurate growth factor for higher return periods.

ii) The regional flood frequency analysis has been carried out by considering more influential variables.

iii) Flood forecasting may be done by considering hourly based hydro-meteorological data, travel time and forecasted data to achieve a better lead time.

iv) Flood inundation modeling may be carried out by using advanced modeling, high spatial resolution Satellite images, by taking data of recent flood year, more cross sectional area data below the delta and also by considering the topography of the region.

CHAPTER 9
REFERENCES

- Aaron Cook, V. M. (2009). "Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping". Journal of Hydrology, 377,131–142.
- Adamowski, K. (2000). "Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods". Journal of Hydrology, 229(3–4), 219–231.
- Adnan, N.A., and Atkinson, P.M. (2012). "Remote Sensing of River Bathymetry for Use in Hydraulic Model Prediction of Flood Inundation". IEEE 8th International Colloquium on Signal Processing and its Applications, 159-163.
- American Society of Civil Engineers. (1949). "Hydrology Handbook." Committee on Hydrology Issue 28 of Manuals of Engineering Practice, American Society of Civil Engineers.
- Anderson D.J. (2000). "GIS-based hydrologic and hydraulic modeling for flood delineation at highway river crossings". Unpublished Master's Thesis, Univ. of Texas at Austin, Austin, Texas, USA.
- Aqil, M., Kita, I., Yano, A., and Nishiyama, S. (2007). "A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff". Journal of Hydrology, 337, 22–34.
- Ashkar, F., and Rousselle, J. (1983). "Some remarks on the Truncation used in partial flood series models". Water Resources Research, 19(2), 477–480.
- Asokan, S.M., and Dutta, D. (2008). "Analysis of water resources in the Mahanadi River Basin, India under projected climate conditions". Hydrological Process, 22, 3589–3603.
- Baeriswyl, P.A., and Rebetez, M. (1997). "Regionalization of precipitation in Switzerland by means of principal component analysis". Theoretical and Applied Climatology, 58(1-2), 31-41.
- Bashir, A., Muhammad, S.K., Butt, M.J., and Dahri, Z.H. (2010). "Hydrological modelling and flood hazard mapping of Nullah Lai". Proceedings Pakistan Academy of Sciences, 47(4), 215–226.
- Bates, P.D., Anderson, M.G., Hervouet, J.M., and Hawkes, J.C. (1997). "Investigating the behaviour of two-dimensional finite element models of compound channel flow". Earth Surface Processes and Landforms 22(1), 3–17.

- Beaulieu, C., Seidou, O., Ourada, T. B. M. J., and Zhang, X. (2009). "Intercomparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change point models". Water Resources Research, 45(8), 15.
- Becker, A., and Grunewald, U. (2003). "Flood risk in Central Europe". Science, 300, 1099-1099.
- Begueria, S. (2005). "Uncertainties in partial duration series modeling of extremes related to the choice of threshold value". Journal of Hydrology, 303(1–3), 215–230.
- Benson, M. A. (1968). "Uniform flood-frequency methods for federal agencies". Water Resources Research, 4(5), 891–908.
- Birikundavyi, S., Rouselle, J., and Nguyen, V.T.V. (1997). "Estimation regionale de quantiles de crues par l'analyse des correspondances". Canadiaen Journal of Civil Engineering, 24(3), 438–447.
- Birsan, M.V., Molnar, P., Burlando, P., and faundler, M. (2005). "Streamflow trends in Switzerland". Journal of Hydrology, 314 (1–4), 312–329.
- Bhatt, C.M., Rao, G.S., Manjushree, P., and Bhanumurthy, V. (2010). "Space Based Disaster Management of 2008 Kosi Floods, North Bihar, India". Journal of Indian Society of Remote Sensing, 38 (2010), 99-108.
- Bobee, B., and Robitaille, R. (1975). "The use of the Pearson type 3 distribution and log Pearson type 3 distribution revisited". Water Resources Research, 13(2), 427–443.
- Bobee, B., Cavadias, G., Ashkar, F., Bernier, J., and Rasmussen, P. (1993). "Towards a systematic approach to comparing distributions used in flood frequency analysis". Journal of Hydrology, 142(1–4), 121–136.
- Bocchiola, D., De Michele, C., and Rosso, R. (2003). "Review of recent advances in index flood Estimation". Hydrologial Earth System Science, 7(3), 283–296.
- Booij, M.J. (2005). "Impact of climate change on river flooding assessed with different spatial model resolutions". Journal of Hydrology, 303, 176– 198.
- Borujeni,S.C., Sulaiman, W.N.A., and Eslamian, S. (2010). "Regional flood frequency analysis using L-moments for North Karoon basin, Iran". Journal of Flood Engineering, 1(1), 67–76.
- Brath, A., Castellarin, A., Franchini, M., and Galeati, G. (2001). "Estimating the index flood using indirect methods". Hydrological Sciences Journal, 46(3), 399-418.
- Buishand, T. A. (1989). "The PDS method with a fixed number of peaks". Journal of Hydrology, 109(1–2), 1–9.
- Burn, D. H. (1989). "Cluster analysis as applied to regional flood frequency". Journal of Water Resources Planning and Management, 115(5), 567–582.

- Burn, D.H. (1990). “Evaluation of regional flood frequency-analysis with a region of influence approach”. Water Resources Research, 26(10), 2257-2265.
- Burn, D.H. (1997). “Catchment similarity for regional flood frequency analysis using seasonality measures”. Journal of Hydrology, 202(1-4), 212-230.
- Burn, D. H., and Goel, N. K. (2000). “The formation of group for regional flood frequency analysis”. Hydrological Sciences, 45(1), 97–112.
- Campolo, M., Andreeussi, P., and Soldati, A. (1999). “River Flood Forecasting with a Neural Network Model”. Water Resources Research, 35 (4), 1191-1197.
- Campolo, M., Soldati, A. and Andreeussi, P. (2003). “Artificial Neural Network Approach to Flood Forecasting in the River Arno”. Hydrological Sciences Journal, 48 (4), 381-398.
- Castellarin, A., Burn, D. H., and Brath, A. (2001). “Assessing the effectiveness of hydrological similarity measures for regional flood frequency analysis”. Journal of Hydrology, 241(3–4), 270–285.
- Central Water Commission (CWC). (1997). “Report of the Working Group on Flood Management for the Ninth Fiver Year Plan (1997–2002)”. Central Water Commission, New Delhi, India.
- Chiang, S.M., Tsay, T.K., and Nix, S.J. (2002a). “Hydrologic regionalization of watersheds. I:Methodology development”. Journal of Water Resources Planning and Management, 128(1), 3-11.
- Chiang, S.M., Tsay, T.K., Nix, S.J. (2002b). “Hydrologic regionalization of watersheds. II: Applications”. Journal of Water Resources Planning and Management, 128(1), 12-20.
- Chang, T.J., Hsu, M.K., and Huang, C.J. (2000). “A GIS distributed watershed model for simulating flooding and inundation”. Journal of the American Water Resources Association, 36, 975-988.
- Chang, F. J., and Chen, Y. C. (2001). “A counter propagation fuzzy-neural network modeling approach to real time stream flow prediction”. Journal of Hydrology, 245, 153–164.
- Chang, F. J., Chang, L. C., and Huang, H. L. (2002). “Real-time recurrent neural network for stream-flow forecasting”. Hydrological Processes, 16, 2577–2588.
- Chang, F. J., and Chang, Y. T. (2006). “Adaptive neuro-fuzzy inference system for prediction of water level in reservoir”. Advance Water Resources, 29, 1–10.
- Chang, F. J., Chiang, Y. M., and Chang, L.C. (2007). “Multi-step-ahead neural networks for flood forecasting”. Hydrological Sciences Journal, 521, 114–130.

- Chau, K. W., Wu, C. L., and Li, Y. S. (2005). "Comparison of several flood forecasting models in Yangtze River". Journal of Hydrological Engineering, ASCE 10(6), 485–491.
- Chavoshi, S., and Soleiman, W.N.A. (2009). "Delineating pooling group for flood frequency analysis using soft computing". European Journal of Science and Research, 35(2), 181–187.
- Chen, Y., Cuddy, S.M., Wang, B., Merrin, L.E., Pollock, D., and Sims, N. (2011). "Linking inundation timing and extent to ecological response models using the Murray-Darling Basin Floodplain Inundation Model (MDBFIM)". In: Chan F, Marinova D, Anderssen RS (eds) Proceedings of MODSIM2011: The 19th International Congress on Modelling and Simulation: 4092-4098.
- Chen, Y., Wang, W., Cuddy, S., Pollino, C., and Merrin, L.E. (2012). "Spatial modelling of potential water retention under floodplain inundation using remote sensing and GIS". In: Proceedings of iEMSs 2012: The 6th International Congress on Environmental Modelling and Software Modelling, Leipzig, Germany.
- Chen, Y., Huang, C., Ticehurst, C., Merrin, L. and Thew, P. (2013). "An evaluation of MODIS daily and 8-day composite products for floodplain and wetland inundation mapping". Wetlands. DOI 10.1007/s13157-013- 0439-4.
- Chow, V.T., Maidment, D.R., and Mays, L.W. (1988). "Applied hydrology". McGraw-Hill, New York NY, USA.
- Chowdhary, A.K. (2005). "Regional flood frequency analysis: A case study of Mahanadi basin". Proceeding of Global Conf. on Flexible Systems Management, Global Institute of Flexible Systems Management, Bhopal, India, 456–461.
- Claps, P., and Laio, F. (2003). "Can continuous streamflow data support flood frequency analysis? An alternative to the partial duration series approach". Water Resources Research, 39(8), 1–11.
- Coulibaly, P., Anctil, F., and Bobee, B. (2000). "Daily Reservoir Forecasting Using Artificial Neural Network with Stopped Training Approach". Journal of Hydrology, 230(3-4), 244-257.
- Cretaux, J.F., Nguyen, M.B., Delclaux, F., Mognard, N., Lion, C., Pandey, R.K., Tweed, S., Calmant, S., and Maisongrande, P. (2011). "Flood mapping inferred from remote sensing data". Fifteenth International Water Technology Conference, Alexandria, Egypt. 1:48-62.
- Cunderlik, J.M., and Burn, D.H. (2002). "The use of flood regime information in regional flood frequency analysis". Hydrological Sciences Journal, 47(1), 77-92.
- Cunderlik, J.M., and Burn, D.H. (2006b). "Site-focused nonparametric test of regional homogeneity based on Flood regime". Journal of Hydrology, 318(1-4), 301-315.

- Cunderlik, J.M., and Ouarda, T.B.M.J. (2006). "Regional flood-duration-frequency modeling in the changing environment". Journal of Hydrology, 318(1-4), 276-291.
- Cunnane, C. (1973). "A particular comparison of annual maximum and partial duration series methods of flood frequency prediction". Journal of Hydrology, 18(3–4), 257–271.
- Cunnane, C. (1979). "A note on the Poisson assumption in partial duration series models". Water Resources Research, 15(2), 489–494.
- Cunnane, C. (1985). "Factors affecting choice of distribution for flood series". Hydrological Sciences Journal, 30(1), 25–36.
- Cunnane, C. (1988). "Methods and merits of regional flood frequency-analysis". Journal of Hydrology, 100(1-3), 269-290.
- Cunnane, C. (1989). "Statistical distributions for flood frequency analysis". Volume WMO Publ. No. 718. World Meteorological Organization, Geneva, Switzerland Operational Hydrology Report No. 33rd Edition.
- Dalrymple, T. (1960). "Flood-frequency analysis". Water Supply Paper 1543-A, USGS, Washington, DC, USA.
- Das, S., and Cunnane, C. (2010). "Examination of homogeneity of selected Irish pooling groups". Hydrological Earth System Science, 7, 5099–5130.
- Dawson, C. W., and Wilby, R. L. (1998). "An artificial neural network approach to rainfall runoff modeling". Hydrological Sciences Journal, 43(1), 47–67.
- Dawson, C. W., and Wilby, R. L. (2001). "Hydrological Modeling Using ANN". Progress in Physical Geography, 25(1), 80-108.
- Deidda, R., and Puliga, M. (2009). "Performances of some parameter estimators of the PP distribution over rounded-off samples". Physics and Chemistry of the Earth, 34, 626–634.
- Deidda, R. (2010). "A multiple threshold method for fitting the generalized Pareto distribution and a simple representation of the rainfall process". Hydrological Earth System Science, Discuss, 7, 4957–4994.
- De Michele, C., and Rosso, R. (2001). "Uncertainty assessment of regionalized flood frequency estimates". Journal of Hydrological Engineering, 6(6), 453–459.
- Demuth, H. B., and Beale, M. (1998). "Neural Network Toolbox for Use with MATLAB, Users Guide". Mathworks Inc., Natick Massachusetts, USA.
- Dolling, O. R., and Vears, E. A. (2002). "Artificial neural network for stream flow prediction". Journal of Hydraulics Research, 40, 547–554.
- Douglas, E.M., Vogel, R.M., and Kroll, C.N. (2000). "Trends in floods and low flows in the United States: impact of spatial correlation". Journal of Hydrology, 240, 90–105.

- Dufek, A.S., and Ambrizzi, T. (2008). "Precipitation variability in São Paulo State, Brazil". Theory of Applied Climatology, 93 (3), 167–178.
- Eng, K., Tasker, G. D., and Milly, P. C. D. (2005). "An analysis of region of- influence methods for flood regionalization in the Gulf-Atlantic rolling plains". Journal of the American Water Resources Association, 41(1), 135–143.
- Eng, K., Stedinger, J. R., and Gruber, A. M. (2007). "Regionalization of streamflow characteristics for the Gulf-Atlantic rolling plains using leverage-guided region-of-influence regression". World Environmental and Water Resources Conf., Restoring our Natural Habitat, Tampa, FL, K. C. Kabbes, ed., ASCE, Reston, VA, USA.
- Eslamian,S.S., and Biabanaki, M.(2008). "Low flow regionalization modeling". International Journal of Ecological Economics and Statistics, 12(F08), 82–97.
- Fausett, L. (1994). "Fundamentals of Neural Networks Architectures, Algorithms, and Applications". Prentice-Hall, Upper Saddle River, New Jersey, USA.
- FEH. (1999). "Flood Estimation Handbook (FEH), volume 3 of Statistical procedures for flood frequency estimation". Institute of Hydrology, Wallingford, UK.
- Fill, H. D., and Stedinger, J. R. (1998). "Using regional regression within index flood procedures and an empirical Bayesian estimator". Journal of Hydrology, 210(1–4), 128–145.
- Fletcher, R. (1987). "Practical Methods of Optimization". John Wiley and Sons, New York, NY, USA.
- Flood workshop, (2011). Workshop on floods in Odisha: civil society perspectives. 16th October 2011, Conference Hall, Odisha Environmental Society, Bhubaneswar, Odisha, India.
- Foster, H. A. (1924). "Theoretical frequency curves". ASCE Trans., 87, 142–203.
- Fovell, R. G., and Fovell, M. Y. C. (1993). "Climate zones of the conterminous United States using cluster analysis". Journal of Climate, 6(11), 2103–2135.
- FSR (1975). "Flood Studies Report (FSR), volume 1". National Environment Research Council, London, UK.
- Fuller, W. E. (1914). "Flood flows". ASCE Trans., 77, 567–617.
- Gaal, L., Kysely, J., and Szolgay, J. (2008). "Region-of-influence approach to a frequency analysis of heavy precipitation in slovakia". Hydrology and Earth System Sciences, 12(3), 825-839.
- Gao, B.C. (1996). "NDWI- A normalized difference water index for remote sensing of vegetation liquid water from space". Remote Sensing of Environment 58(3) 257- 266.
- Gebeyehu, A. (1989). "Regional flood frequency analysis". Unpublished Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden.

- Gebregiorgis, A. S., Moges, S.A., and Awulachew, S.B. (2013). "Basin Regionalization for the Purpose of Water Resource Development in a Limited Data Situation: Case of Blue Nile River Basin, Ethiopia". Journal of Hydrologic Engineering, 18 (10), 1349-1359.
- Gee, D.M., Anderson, M.G., and Baird, L. (1990). "Large scale floodplain modelling". Earth Surface Processes and Landforms, 15, 512-523.
- Ghosh, S., Raje, D., and Mujumdar, P.P. (2010). "Mahanadi stream flow: climate change impact assessment and adaptive strategies". Current Science, 98, 1084–1091.
- Goel, N.K., Kurothe, R.S., Mathur, B.S., and Vogel, R.M. (2000). "A derived flood frequency distribution for correlated rainfall intensity and duration". Journal of Hydrology, 228, 56-67.
- Goel, N.K., Than, H.H., Arya, D.S. (2005). "Flood Hazard Mapping in the Lower Part of Chindwin River Basin, Myanmar". International Proceedings of the International Conference on Innovation Advances and Implementation of Flood Forecasting Technology, Tromsø, Norway, 17–19 October.
- Goldberg, D. E. (1989). "Flood routing with neural networks using genetic algorithm". Proceedings, Engineering Hydrology, ASCE, Reston, Va., USA, 754–759.
- Gordon, N. D., McMahon, T. A., Finlayson, B. L., Gippel, C. J., and Nathan, R. J. (2004). "Stream hydrology: An introduction for ecologists", Wiley, Chichester, West Sussex, UK.
- Gosain, A.K., Rao, S., Basuray, D. (2006). "Climate change impact assessment on hydrology of Indian River basins". Current Science, 90, 346–353.
- Goswami, B.N., Venugopal, V., Sengupta, D., Madhusoodanan, M.S., and Xavier, P.K. (2006). "Increasing trend of extreme rain events over India in a warming environment". Science 314, 1442–1445.
- Greis, N.P., and Wood, E. F. (1981). "Regional flood frequency estimation and network design". Water Resources Research, 17, 1167-1177.
- Greis, N. P., and Wood, E. F. (1983). "Correction". Water Resources Research, 19(2), 589-590.
- Griffis, V. W., and Stedinger, J. R. (2007a). "Evolution of flood frequency analysis with Bulletin 17". Journal of Hydrological Engineering, 12(3), 283–297.
- Grover, P.L., Burn, D.H., and Cunderlik, J.M. (2002). "A comparison of index flood estimation procedures for ungauged catchments". Canadian Journal of Civil Engineering, 29(5), 734-741.
- Gumbel, E.J. (1941). "The return period of flood flows". The Annals of Mathematical Statistics, 12(2), 163-190.

- Hagan, M. T., Demuth, H. B., and Beale, M. (1996). "Neural Network Design". Cengage Learning, New Delhi, India.
- Haile, AT. (2005). "Integrating hydrodynamic models and high resolution DEM (LIDAR) for flood modeling". MSc thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, the Netherlands.
- Hall, M. J., and Minns, A. W. (1999). "The classification of hydrologically homogeneous regions". Hydrological Sciences Journal, 44(5), 693–704.
- Han, K.Y., Lee, J.T., and Park, J.K. (1998). "Flood inundation analysis resulting from levee break". Journal of Hydraulic Research, 36, 747-759.
- Hassanpour, F., Bahraini, M. M., and Amiri, M. (2012). "Flood zoning using Hec-GeoRAS". Zeitune Sabz Press, Tehran.
- Haykin, S. (1999). "Neural Networks: A Comprehensive Foundation". Macmillan, New York, USA.
- Hazen, A. (1914). "Discussion of 'Flood flows' by W. E. Fuller". ASCE Trans., 77, 626–632.
- Hazen, A. (1921). "Discussion of 'The probable variations in yearly runoff as determined from a study of California streams' by L. Standish Hall". ASCE Trans., 84, 214–222.
- Hosking, J. R. M., and Wallis, J. R. (1987). "Parameter and quantile estimation for the generalized Pareto distribution." Technometrics, 29(3), 339–349.
- Hosking, J. R. M., and Wallis, J. R. (1988). "The effect of intersite dependence on regional flood frequency analysis". Water Resources Research, 24(4), 588–600.
- Hosking, J. R. M., and Wallis, J. R. (1993). "Some statistics useful in regional frequency analysis." Water Resour. Res., 29(2), 271–281.
- Hosking, J. R. M., and Wallis, J. R. (1997). "Regional frequency analysis: An approach based on L-moments". Cambridge University Press, New York, NY, USA.
- Hoyt, W. G. (1936). "Studies of relations of rainfall and runoff in the United States". Water Supply Paper 772, USGS, Washington, DC, USA.
- Huang, C., Chen, Y., Wu, J., and Yu, J. (2012a). "Detecting floodplain inundation frequency using MODIS timeseries imagery". In: Proceedings of Agro-Geoinformatics2012: The 1st International Conference on Agro- Geoinformatics, Shanghai, China.
- Huang, S., Li, J., and Xu, M. (2012b). "Water surface variations monitoring and flood hazard analysis in Donting Lake area using long-term Terra/MODIS data time series". Natural Hazards, 62, 93-100.

- Hundecha, Y., Bardossy, A., and Theisen, H.W. (2001). "Development of a fuzzy logic based rainfall–runoff model". Hydrological Sciences Journal, 46(3), 363–377.
- Hydrologic Engineering Center. (2002). "HEC-RAS User's Manual". U.S. Army Corps of Engineers, Waterways Experimentation, Vicksburg, MS, USA.
- Irvine, K.N., and Waylen, P.R. (1986). "Partial series analysis of high flows in Canadian rivers". Canadian Water Resources Journal 11 (20), 83–91.
- Iwasa, Y., and Inoue, K. (1982). "Mathematical simulations of channel and overland flood flows in view of flood disaster engineering". Journal of Natural Disaster Science, 4 (1), 1-30.
- Jain, A., and Srinivasulu, S. (2004). "Development of effective and efficient rainfall-runoff models using integration of deterministic, realcoded genetic algorithms and artificial neural network techniques". Water Resources Research, 404, W04302.
- Jaiswal, R. K., Goel, N. K., Singh, P., and Thomas, T. (2003). "L-Moment based Flood FrequencyModelling". IE (I) Journal -CV, 84, 6-10.
- Jakob, D., Reed, D. W., and Robson, A. J. (1999). "Selecting a pooling group (B)". Flood estimation handbook, Institute of Hydrology, Wallingford, UK, 3(26).
- Jang, J.S. R., Sun, C. T., and Mizutani, E. (1997). "Neuro-fuzzy and Soft Computing". Prentice-Hall, Upper Saddle River, New Jersey, USA.
- Jarvis, C. S. (1926). "Flood flow characteristics". ASCE Trans., 89, 985–1033; closure, 1091-1104.
- Jayasuriya, M.D.A., and Mein, R.G. (1985). "Frequency Analysis Using the Partial Series". Hydrology and Water Resources Symposium, IEAust. Natl. Conf. Publ. No. 85/2, 81-85.
- Jena, P. P., Chatterjee, C., Pradhan, G., and Mishra, A. (2014). "Are recent frequent high floods in Mahanadi basin in eastern India due to increase in extreme rainfalls?". Journal of Hydrology, 517, 847-862.
- Jingyi, Z., and Hall, M.J. (2004). "Regional flood frequency analysis for the Gan Ming river basin in China". Journal of Hydrology, 296, 98–117.
- Kaiser, H. F. (1960). "The application of electronic computers to factor analysis". Educational and Psychological Measurement, 20(1), 141–151.
- Kar, A.K., and Lohani, A.K. (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.
- Kar, A.K., Winn, L.L., Lohani, A.K., and Goel, N.K. (2012). "Application of Clustering Techniques Using Prioritized Variables in Regional Flood Frequency Analysis-Case Study of Mahanadi Basin". Journal of Hydrologic Engineering, 17(1), 213-223.

- Kar, A.K., Winn, L.L., Lohani, A.K., and Goel, N.K. (2012). "Soft Computing–Based Workable Flood Forecasting Model for Ayeyarwady River Basin of Myanmar". Journal of Hydrologic Engineering, 17(7), 807-822.
- Karunanithi, N., Grenney, W. J., and Whitley, D. (1994). "Neural network for river flow prediction". Journal of Computing in Civil Engineering 8, 201–220.
- Kendall, M.G. (1975). "Rank Correlation Methods". Griffin, London, USA.
- Kinnison, H. B. (1930). "The New England flood of November, 1927". Water Supply Paper 636-C, USGS, Washington, DC, USA.
- Kinnison, H. B., and Colby, B. R. (1945). "Flood formulas based on drainage-basin characteristics". ASCE Trans., 110, 849–904.
- Khatua, K.K., and Mahakul, B. (1999). "Flood in Mahanadi delta stage II area – a case study". International Proceeding of National Seminar on Disaster Management, 12–13 November, 1999, University College of Engineering Burla, Odisha, India.
- Khatua, K.K., Patra, K.C. (2004). "Management of high flood in Mahanadi and its tributaries below Naraj". 49th Annual session of IEI (India) 2nd February 2004, Odisha state center, Bhubaneswar, Odisha, India.
- Knebl, M.R., Yang, Z.L., Hutchison, K., and Maidment, D.R. (2005). "Regional scale flood modeling using NEXRAD rainfall, GIS, and HEC-HMS/RAS: A case study for the San Antonio River Basin summer 2002 storm event". Journal of Environmental Management, 75(4), 325-336.
- Konecny, F., and Nachtnebel, H.P. (1985). "Extreme Value process and the evaluation of risk in flood analysis". Appl. Mathematical Modelling 9, 11–15.
- Kumar, R., Singh, R. D., and Seth, S. M. (1999). "Regional flood formula for seven subzones of India". Journal of Hydrological Engineering, 4(3), 240–244.
- Kumar, R., and Chatterjee, C. (2005). "Regional flood frequency analysis using L-moments for North Brahmaputra Region of India". Journal of Hydrological Engineering, 10(1), 1–7.
- Kumar, S., Merwade, V., Kam, J., and Thurner, K. (2009). "Stream flow trends in Indiana: effects of long term persistence, precipitation and subsurface drains". Journal of Hydrology. 374, 171–183.
- Langa, M., Quardab, T. B. M. J., and Bobee, B. (1999). "Towards operational guidelines for over-threshold modeling". Journal of Hydrology, 225(3–4), 103–117.
- Legates, D.R., and McCabe, G.J. (1999). "Evaluating the use of "goodnessof-fit" measures in hydrologic and hydroclimatic model validation". Water Resources Research 35(1), 233–241.

- Lekkas, D. F., Imrie, C. E., and Lees, M. J. (2001). "Improved non-linear transfer function and neural network methods of flow routing for real-time forecasting". Journal of Hydroinformatics, 3,153–164.
- Lettenmaier, D., and Potter, K. W. (1985). "Testing flood frequency estimation methods using a regional flood generation model". Water Resources Research, 21(12), 1903–1914.
- Lettenmaier, D.P., Wallis, J.R., and Wood, E.F. (1987). "Effect of regional hereogeneity on flood frequency estimation". Water Resources Research, 23(2), 313–324.
- Lettenmaier, D. P., and Wood, E. F. (1993). "Hydrologic forecasting". Handbook of Hydrology, Devid R. Maidment, ed., McGraw-Hill, New York, NY, USA.
- Lim, Y. H., and Lye, L. M. (2003). "Regional flood estimation for ungauged basins in Sarawak, Malaysia". Hydrological Sciences Journal, 48(1), 79–94.
- Lohani, A.K., Goel, N.K., and Bhatia, K.K.S. (2006). "Takagi–Sugeno fuzzy inference system for modeling stage–discharge relationship". Journal of Hydrology, 331,146–160.
- Lopez, A.G. (2004). "Hydrological forecasting using the double index flood method". British Hydrological Society.
- Lu, L. H., and Stedinger, J. R. (1992). "Sampling variance of normalized gev pwm quantile estimators and a regional homogeneity test". Journal of Hydrology, 138(1-2), 223-245.
- Madsen, H., Rosbjerg, D., and Harremoes, P. (1994). "PDS-modelling and regional Bayesian estimation of extreme rainfalls". Nord Hydrology, 25(4), 279–300.
- Madsen, H., Rasmussen, P. F., and Rosbjerg, D. (1997). "Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events, 1. At-site modeling". Water Resources Research, 33(4), 747–757.
- Malekinezhad, H., Nachtnebel, H. P., and Klik, A. (2011). "Comparing the index-flood and multiple-regression method using L-moments". Physics and Chemistry of Earth, 36(1–4), 54–60.
- Martins, E. S., and Stedinger, J.R. (2001). "Generalized Maximum Likelihood Pareto-Poisson Flood Risk Analysis for Partial Duration Series". Water Resources Research, 37(10), 2559-2567.
- McDermott, G.E, and Pilgrim, D.H. (1982). "Design Flood Estimation for Small Catchments in New SouthWales". Australian Water Resources Council Technical Paper No. 73, Department of National Development and Energy, 233.

- Meng, F., Li, J., and Gao, L. (2007). “ERM-POT model for quantifying operational risk for Chinese commercial banks”. Lecture notes in computer science, Y. Shi, ed., Springer, Berlin, Heildelberg, Germany.
- Merz, R., and Bloschl, G. (2004). “Flood frequency regionalisation-spatial proximity vs. catchment attributes”. Journal of Hydrology, 302(1-4), 283-306.
- Minns, A. W., and Hall, M. J. (1996). “Artificial Neural Networks as Rainfall-Runoff Models”. Hydrological Sciences Journal, 41 (3), 399-417.
- Mishra, P.K., and Behera, S. (2009). “Development and management of water and energy resource”. 7th International R&D Conference, 4–6 February, Bhubaneswar, Odisha, India.
- Mosley, M. P. (1981). “Delimitation of New Zealand hydrologic regions”. Journal of Hydrology, 49(1), 173–192.
- Mukerjee, A., Chatterji, C., Singh, N., and Raghuvanshi, N. S. (2009). “Flood Forecasting Using ANN, Neuro-Fuzzy and Neuro-GA Models”. Journal of Hydrologic Engineering, ASCE, 14(6), 647-652.
- Nathan, R. J., and McMahon, T. A. (1990). “Identification of homogenous regions for the purpose of regionalization”. Journal of Hydrology, 121(1–4), 217–238.
- National Institute of Hydrology. (1985). “Design Flood Estimation for Narmada Sagar and Sardar Sarovar Project”. Final Report 1, July 1985, Roorkee, India.
- National Institute of Hydrology. (1997). “Regional flood frequency analysis using L-moment”. Technical Rep. TR (BR), Roorkee, India.
- Nayak, P. C., Sudheer, K. P., Rangan, D. M., and Ramasastri, K. S. (2004). “A neuro-fuzzy computing technique for modeling hydrological time series”. Journal of Hydrology, 2911-2, 52–66.
- Nayak, P. C., Sudheer, K. P., and Ramasastri, K. S. (2005a). “Fuzzy computing based rainfall-runoff model for real time flood forecasting”. Hydrological Processes, 194, 955–968.
- Nayak, P. C., Sudheer, K. P., Rangan, D. M., and Ramasastri, K. S. (2005b). “Short-term flood forecasting with a neurofuzzy model”. Water Resources Research, 41, W04004.
- Neal, J. (2012). "A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas". Water Resources Research 48(11).
- Norbiato, D., Borga, M., Sangati, M., and Zanon, F. (2007). “Regional frequency analysis of extreme precipitation in the eastern Italian Alps and the August 29, 2003 flash flood”. Journal of Hydrology, 345(3–4), 149–166.

- O'Connell, P. P. L. (1868). "On the relation of the freshwater floods of rivers to the areas and physical features of their basins and on a method of classifying rivers and streams with reference to the magnitude of their floods". Minutes Proc. Inst. Civ. Eng., 27(1868), 204–217.
- Onoz, B., and Bayazit, M. (2001). "Effect of the occurrence process of the peaks over threshold on the flood estimates". Journal of Hydrology, 244(1–2), 86–96.
- Ordoyne, C., and Friedl, M.A. (2008). "Using MODIS data to characterize seasonal inundation patterns in the Florida Everglades". Remote Sensing of Environment, 112(11), 4107-4119.
- Ormsby, J.P., Blanchard, B.J., and Blanchard, A.J. (1985). "Detection of lowland flooding using active microwave systems". Photogrammetric Engineering and Remote Sensing, .51(3), 317-28.
- Orok, H. (2011). "A GIS Based Flood Risk Mapping of Kano City, Nigeria". University of East Anglia, Norwich, UK.
- Pal, I., and Al-Tabbaa, A. (2011a). "Assessing seasonal precipitation trends in India using parametric and non-parametric statistical techniques". Theory of Applied Climatology, 103, 1–11.
- Pal, I., and Al-Tabbaa, A. (2011b). "Monsoon rainfall extreme indices and tendencies from 1954–2003 in Kerala, India". Climatic Change 106, 407–419.
- Parhi, P.K., Mishra, S.K., Singh, R., and Tripathy, V.K. (2012). "Floods in Mahanadi river basin, Orissa (India): a critical review. India Water Week". Water, Energy and Food Security: Call for Solutions, 10–14 April, New Delhi.
- Parida, B. P. (2004). "A partitioning methodology for identification of homogeneous regions in regional flood frequency analysis". Proceedings of BALWOIS Conference, Ohrid, Republic of Macedonia.
- Patro, S., Chatterjee, C., Mohanty, S., Singh, R., and Raghuwanshi, N.S. (2009). "Flood Inundation Modeling using MIKE FLOOD and Remote Sensing Data". Journal of Indian Society of Remote Sensing, 37, 107-118.
- Patterson, D. (1996). "Artificial Neural Networks". Prentice-Hall, Singapore.
- Pearson, C.P. (1991). "New Zealand regional flood frequency analysis using L-moments, volume 30". New Zealand Hydrological Society, Christchurch, Nouvelle-Zelande, Journal of Hydrology, New Zealand.
- Pearson, C. P., and Henderson, R. D. (1998). "Frequency distributions of annual maximum storm rainfalls in New Zealand". Journal of Hydrology, 37(1), 19–33.
- Pham, H. X., Asaad, Y. S., and Bruce, M. (2014). "Statistical properties of partial duration series: Case study of North Island, New Zealand". Journal of Hydrological Engineering, 807–815.

- Pham, H. X., Asaad, Y. S., and Bruce, M. (2014). "Statistical Properties of Partial Duration Series and Its Implication on Regional Frequency Analysis". Journal of Hydrological Engineering, 19 (7), 1471-1480.
- Pultz, T.J., Crevier, Y., Brown, R.J., and Boisvert, J. (1997). "Monitoring of local environmental conditions with SIR-C/XSAR". Remote Sensing of the Environment, 59(4), 248-255.
- Raje, D., and Mujumdar, P.P. (2009). "A conditional random field-based downscaling method for assessment of climate change impact on multisite daily precipitation in the Mahanadi basin". Water Resources Research, 45, W10404.
- Rakesh, K., and Singh, R.D.(1994). "Development of regional flood formula for Mahanadi subzone- 3d". TR (BR) 134, Report of NIH, Roorkee.
- Rao, P.G. (1993). "Climatic changes and trends over a major river basin in India". Climate Research, 2, 215–223.
- Rao, P.G. (1995). "Effect of climate change on stream flows in the Mahanadi river basin, India". Water International, 20, 205–212.
- Rao, A. R., and Hamed, K. H. (2000). "Flood frequency analysis". CRC Press, Boca Raton, FL,USA.
- Rao, A. R., and Srinivas, V. V. (2006). "Regionalisation of watersheds by fuzzy cluster analysis". Journal of Hydrology, 318(1–4), 57–79.
- Rasmussen, P.F., Ashkar, F., Rosbjerg, D., Bobee, B. (1994). "The POT method for flood estimation: a review". Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Extreme Values: Floods and Droughts, Kluwer Academic, Dordrecht, pp. 15–26 vol. 1: Water Science and Technology Library.
- Robayo, O., Whiteaker, T., and Maidment, D. (2004). "Converting a NEXRAD map to a floodplain map". Proceedings of AWRA GIS and Water Resources III Conf., Nashville, Tenn, USA.
- Robson, A., and Reed, D. (1999). "Statistical procedure for flood frequency estimation". Flood estimation handbook, Centre for Ecology & Hydrology, Wallingford, UK.
- Rosbjerg, D., and Madsen, H. (1992). "On the choice of threshold level in partial duration series". In: Østrem, G. (Ed.), Nordic Hydrological Conference, Alta. NHP Rep. 30, Oslo, Norway, 604–615.
- Rosbjerg, D., and Madsen, H. (2004). "Advanced approaches in PDS/POT modeling of extreme hydrological events". Hydrological Sciences Practice, 21st Century, Proceedingd, British Hydrological Society International Conference, B.Webb, N. Arnell, C. Onof, N. MacIntyre, R. Gurney, and C. Kirby, eds., Imperial College, London, UK.

- Roy, L.B., Kumar, S., and Kumar, J.P. (2011). "Flood forecasting system in Kosi Basin". Proceeding Recent Advances in Civil Engineering, ITBHU, 164-169, Varanasi, UP, India.
- Sadiq, I. K., Yang Hong., Jonathan, J., Muhammad, U. K., and Tom De Groeve. (2014). "Multi-Sensor Imaging and Space-Ground Cross-Validation for 2010 Flood along Indus River, Pakistan". Remote Sensing, 6, 2393-2407.
- Saf, B. (2009). "Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey". Water Resources Management, 23(3), 531–551.
- Saf, B. (2010). "Assessment of the effects of discordant sites on regional flood frequency analysis". Journal of Hydrology, 380(3–4), 362–375.
- Samarasinghe, S.M.J.S., Nandalal, H.K., Weliwitiya, D.P., Fowzed, J.S.M., Hazarikad, M.K.,and Samarakoond, L. (2010). "Application of Remote Sensing and GIS for flood risk analysis: a case study at Kalu- Ganga River, Sri Lanka". International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Kyoto Japan, 38(8), 110-115.
- Samuels, P.G. (1985). "Modelling of river and flood plain flow using the finite element method". Hydraulics Research Ltd., Technical Report SR61, Wallingford.
- Seth, S.M., Perumal, M., and Singh, R.D. (1985). "Regional flood frequency analysis". CS-9 Report of NIH, Roorkee, uttarakhand, India.
- Shafiee, Mardiana, Azman Ahmad, and Osman Kadir (2000). "Capability of Radarsat Data In Monsoon Flood Monitoring". Proceedings of GIS development.
- Shamseldin, A. Y., Nasr, A. E. and O'Conner, K. M. (2002). "Comparison of different forms of the multi-layer feed-forward neural network method used for river flow forecasting". Hydological. Earth System Science, 6, 671–684.
- Shamsi., S.U. (2002). "GIS Applications in Floodplain Management". 22nd Annual ESRI International User Conference, July 8–12, San Diego, CA, USA: ESRI.
- Shaohong, S., Chen Pengxiao, C. (2010). "A real-time flood monitoring system based on GIS and hydrological model". 2nd Conference on Environmental Science and Information Application Technology, IEEE, 605-608.
- Shinyie, W. L., and Ismail, N. (2012). "Analysis of T-Year Return Level for Partial Duration Rainfall Series". Sains Malaysiana, 41(11), 1389–1401.
- Shu, C., and Burn, D.H. (2004). "Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement". Journal of Hydrology, 291(1-2), 132-149.
- Singh, R.D., and Seth, S.M. (1985). "Regional flood frequency analysis for Mahanadi basin using Wakeby distribution". Proceeding of Seminar on Flood Frequency Analysis, New Delhi, India.

- Singh, R.D., and Kumar, R. (1991). "Estimation of Discharge Hydrograph for an Ungauged Catchment Using Unit Hydrograph Approach". Proceeding of 4th National Symposium on Hydrology of Minor water Resources Schemes, Madras, TN, India.
- Sivakumar, B., Jayawardena, A. W., and Fernando, T. M. K. G. (2002). "River flow forecasting use of phase-space reconstruction and artificial neural networks approaches". Journal of Hydrology, 265, 225–245.
- Solomatine, D. P., and Xue, Y. (2004). "M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China". Journal of Hydrologic Engineering, 9(6), 491-501.
- Spokkereff, E. (2000). "Flood Forecasting for the River Rhine in the Netherlands". Proceeding of the Symposium the Extreme of Extremes: Extraordinary Floods, Reykjavic, Ireland, July, 347-352.
- Stambuk, A., Stambuk, N., and Konjevoda, P. (2007). "Application of Kohonen self-organising maps (SOM) based clustering for the assessment of religious motivation". 29th International Conference on Information Technology Interfaces, 87–91.
- Stedinger, J. R., Vogel, R. M., and Foufoula, G. E. (1992). "Frequency analysis of extreme events". Chapter 18, Handbook of hydrology, D. Maidment, ed., McGraw Hill, New York, NY, USA.
- Stedinger, J. R., and Lu, L. (1995). "Appraisal of regional and index flood quantile estimators". Stochastic Hydrology Hydraullics, 9(1), 49–75.
- Sudheer, K. P., Gosain, A. K., and Ramasastri, K. S. (2002). "A datadriven algorithm for constructing artificial neural network rainfallrunoff models". Hydrological Processess,166, 1325–1330.
- Sudheer, K. P. (2005). "Knowledge extraction from trained neural network river flow models". Journal of Hydrological Engineering, 104, 264–269.
- Sveinsson, O.G.B., Boes, D.C., and Salas, J.D. (2001). "Population index flood method for regional frequency analysis". Water Resources Research, 37(11), 2733-2748.
- Sveinsson, O. G. B., Salas, J. D., and Boes, D. C. (2003). "Uncertainty of quantile estimators using the population index flood method". Water Resources Research, 39(8), 1206.
- Taesombut, V., and Yevjevich, V. (1978). "Use of partial duration series for estimating the distribution of maximum annual flood peak". Hydrology Paper no. 97, Colorado State University. FortCollins, CO, USA, 71 p.
- Takagi, T., and Sugeno, M. (1985). "Fuzzy identification of systems and its application to modeling and control". IEEE Trans. Syst. Man Cybern., 15(1), 116–132.

- Takeuchi, K. (1984). "Annual maximum series and partial duration series–Evaluation of Langbien's formula and Chow's discussion". Journal of Hydrology, 68(1–4), 275–284.
- Tasker, G. D. (1982). "Comparing methods of hydrologic regionalization". Water Resources Bulletin, 18(6), 965–970.
- Tasker, G. D., and Stedinger, J. R. (1989). "An operational GLS model for hydrologic regression". Journal of Hydrology, 111(1–4), 361–375.
- Tasker, G. D., Hodge, S. A., and Barks, C. S. (1996). "Region of influence regression for estimating the 50-year flood at ungauged sites". Water Resources Bulletin, 32(1), 163–170.
- Tavares, L.V., and da Silva, J.E, (1983). "Partial Series Method Revisited". Journal of Hydrology, 64,1-14.
- Thandaveswara, B. S., and Sajikumar, N. (2000). "Discussion on classification of river basins using artificial neural network". ASCE, Journal of Hydrological Engineering, 5(3), 290–298.
- Thirumalaiah, K., and Deo, M. C. (1998). "River stage forecasting using artificial neural network". Journal of Hydrological Engineering, ASCE 3(1), 26–30.
- Trefry, C. M., Watkins, D. W., and Johnson, D. (2005). "Regional rainfall frequency analysis for the state of Michigan". Journal of Hydrological Engineering, 437–449.
- U.S. Army Corps of Engineers. (2010). "HEC-RAS, User's Manual Version 4.1". Hydrologic Engineering Centre, Davis, CA, USA.
- Valenca, M., and Ludermir, T. (2000). "Monthly stream flow forecasting using an neural fuzzy network model". Proceedings of Sixth Brazilian Symposium on Neural Networks, 117–120. Brazilian Computer Society, Brazil, Washington, DC, USA.
- Van Montfort, M.A.J., and Witter, J.V. (1986)."The Generalized Pareto distribution applied to rainfall depths". Hydrological Sciences Journal, 31 (2), 151–162.
- Viglione, A., Laio, F., and Claps, P. (2007). "A comparison of homogeneity tests for regional frequency analysis". Water Resources Research, 43, 17, 130.
- Vogel, R. M., and Kroll, C. N. (1992). "Regional geohydrologicgeomorphic relationships for estimation of low flow statistics". Water Resources Research, 28(9), 2451–2458.
- Vogel, R.M., and Wilson, I. (1996). "Probability distribution of annual maximum, mean, and minimum stream flows in the United States". Journal of Hydrologic Engineering, 1(2): 69- 76.
- Wallis, J.R. (1980). "Risk and unceratinties in the evaluation of flood events for the design of hydraulic structures". In Pienee Siccita (eds E. Guggino, G. Rossi and E. Todidi), Catania: Fondazione Politecnica del Mediterraneo, 3-36.

- Wang, Q.J. (1996). “Direct Sample Estimators of L-Moments”. Water Resources Research,32(12), 3617-3619.
- Wan Zin, W.Z., Jamaludin, S., Deni, S.M., and Jemain, A.A. (2010). “Recent changes in extreme rainfall events in Peninsular Malaysia: 1971–2005”. Theory of Applied Climatology, 99, 303–314.
- Water Resources Council (WRC). (1976). “Guidelines for determining flood flow frequency”. Bulletin 17 of the Hydrology Subcommittee, Water Resources Council, Washington, DC, USA.
- Water Resources Department, Government of Orissa. (2009). “Time of Flow in Mahanadi River between Hirakud & Mundali”. http://www.dowrorissa.gov.in/Flood/DailyFloodBulletin.html.
- Waylen, P.R., and Woo, M.K. (1983). “Stochastic analysis of high flows in some central British Columbia rivers”. Canadian Journal of Civil Engineering 10 (2), 639–648.
- Weiss, D.J., and Crabtree, R.L. (2011). “Percent surface water estimation for MODIS BRDF 16-day image composites”. Remote Sensing of Environment, 115, 2035-2046.
- Werner, M., Blazkova, S., and Petr, J. (2005). “Spatially distributed observations in constraining inundation modeling uncertainties”. Hydrological Processes 19: 3081–3096.
- Wilks, D. S. (1993). “Comparison of three-parameter probability distributions for representing annual extreme and partial duration precipitation series”. Water Resources Research, 29(10), 3543–3549.
- Willmott, C.J. (1981). “On the validation of models”. Physical Geography 2, 184–194.
- Willmott, C.J. 1984. “On the evaluation of model performance in physical geography”. In Spatial Statistics and Models, Gaile GL, Willmott CJ (eds). D. Reidel: Boston, MA, USA; 443–460.
- Wright, N.G., Villanueva, I., Bates, P.D., Mason, D.C., Wilson, M.D., Pender, G., and Neelz, S. (2008). “Case study of the use of remotely sensed data for modelling flood inundation on the river Severn, U.K”, Journal of Hydraulic Engineering ASCE, 533-540.
- Xiong, L. H., Shamseldin, A. Y., and O’Connor, K. M. (2001). “A nonlinear combination of the forecasts of rainfall–runoff models by the first order Takagi-Sugeno fuzzy system”. Journal of Hydrology, 245, 196–217.
- Yang, T. (2010). “Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China”. Journal of Hydrology, 380(3–4), 386–405.

- Yuguo, D., Bingyan, C., and Zhihong, J. (2008). "A newly discovered GPD-GEV relationship together with comparing their models of extreme precipitation in summer". Advances in Atmospheric Sciences, 25(3), 507–516.
- Zafirakou-Koulouris, A., Vogel, R.M., Craig, S.M., and Habermeier, J. (1998). "L moment diagrams for censored observations". Water Resources Research, 34(5), 1241–1249.
- Zealand, C. M., Burn, D. H., and Simonovic, S. P. (1999). "Short term stream flow forecasting using artificial neural networks". Journal of Hydrology, 214, 32–48.
- Zhang, Z., Dehoff, A.D., and Pody, R.D. (2010). "New approach to identify trend pattern of streamflows". Journal of Hydrological Engineering, 15 (3), 244–248.
- Zheng, N., Tachikawa, Y., and Takara, K. (2008). "A distributed flood inundation model integrating with rainfall-runoff processes using GIS and Remote Sensing data". The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Beijing, China, Vol. XXXVII. Part B4, 1513-1518.
- Zrinji, Z., and Burn, D.H. (1994). "Flood frequency-analysis for ungauged sites using a region of influence approach". Journal of Hydrology, 153(1-4), 1-21.
- Zrinji, Z., and Burn, D.H. (1996). "Regional flood frequency with hierarchical region of influence". Journal of Water Resources Planning and Management, ASCE, 122(4), 245-252.
- Zvi, A. B. (2009). "Rainfall intensity–duration–frequency relationships derived from large partial duration series". Journal of Hydrology, 367(1–2), 104–114.

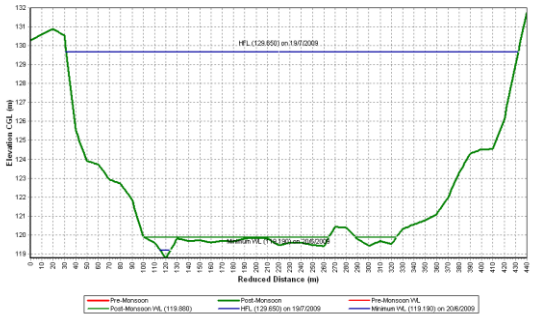
APPENDIX I

List of Discharge Stations

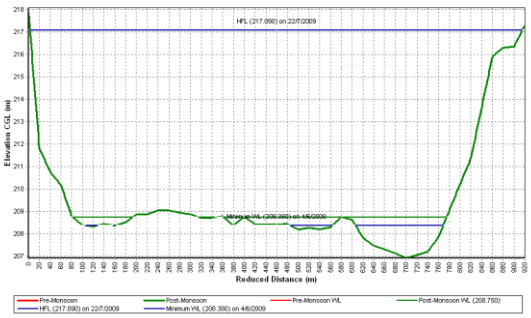
G/D Sites	Latitude (N)	Longitude (E)	Number of years of record
Kantamal	20°39'00	83°43'55"	41
Kesinga	20°11'51"	83°13'30"	34
Sundargarh	22°06'55"	84°00'40"	35
Salebhata	20°59'00"	83°32'22"	39
Tikarapada	20°38'00"	84°37'08"	41
Seorinarayan	21°43'00"	82°35'48"	26
Rajim	20°58'25"	81°52'48"	41
Baronda	20°54'45"	81°53'10"	34
Basantpur	21°43'36"	82°47'17"	41
Jhondhra	21°43'30"	82°20'50"	31
Ghatora	22°03'24"	82°13'15"	31
Kotni	21°14'10"	81°14'50"	32
Kurubhata	21°59'15	83°12'15"	34
Manendragarh	23°12'10"	82°13' 05	22
Pathardihi	21°20'28	81°35'38"	23
Simga	21°37'54"	81°41'16"	40
Rampur	21°39'06"	82°31'10"	40
Sukuma	20°48'30"	84°30'00"	13
Pandigaon	20°05'35"	83°05'00"	9
Paramanpur	21°15'51"	84°16'35"	10
Naraj	20°28'00"	85°42'00"	10
Andhiyarkore	21°50'02"	81°36'21"	34
Bamnidhi	21°53'55"	82°43'02"	41

List of Rainfall Stations

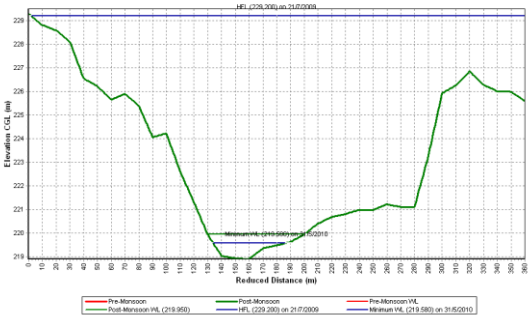
Rainfall Sites	Longitude(E)	Latitude (N)	Number of years of record
BADAPANDUSAR	85.18	20.16	10
BAGHUPALLI	83.88	21.19	10
BANSAJAL	84.25	21.11	10
BARGAON	83.33	20.41	10
CHATIKUDA	83.27	19.97	10
GORLA	83.58	20.61	10
ICHHAPUR	82.63	20.6	10
MADHUPUR	84.83	20.31	10
MAGURBEDA	83.38	20.75	10
NARAJ	85.76	20.47	10
PATORA	82.46	20.74	10
SAGADA	84.07	20.69	10
SAGJURI	84.06	21.05	10



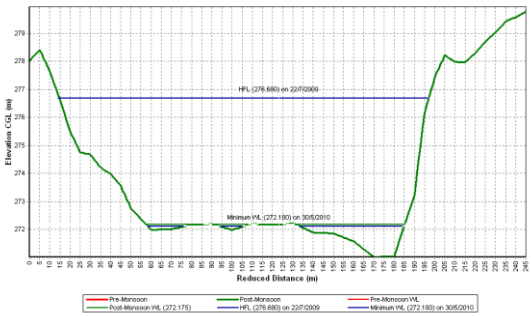
Kantamal



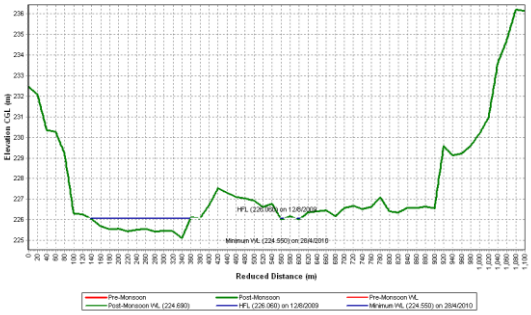
Basantpur



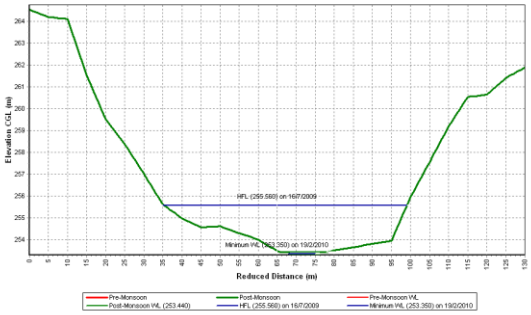
Rampur



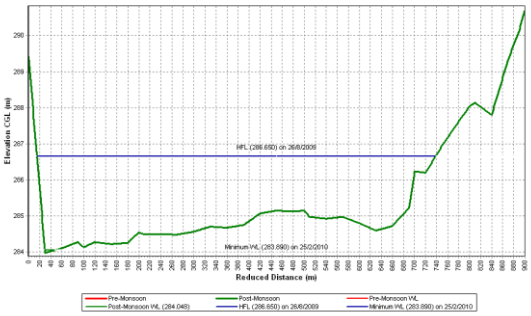
Patherdihi



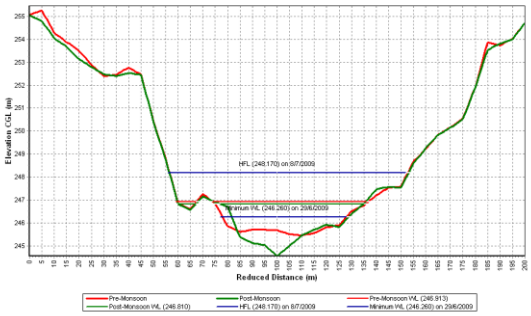
Bamnidhi



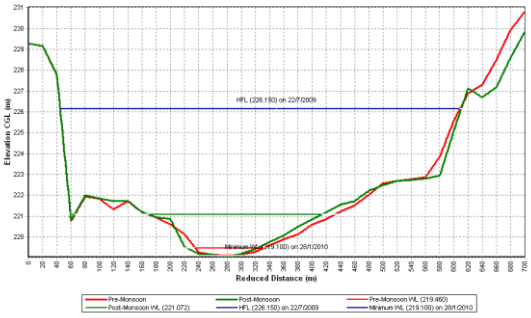
Andhiyarkore



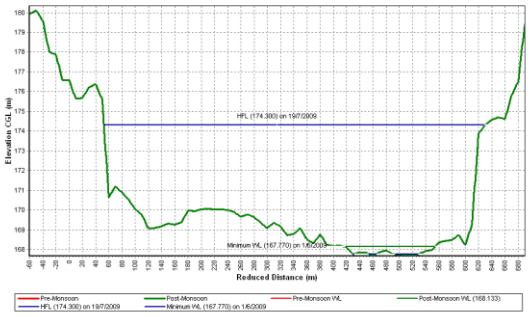
Baronda



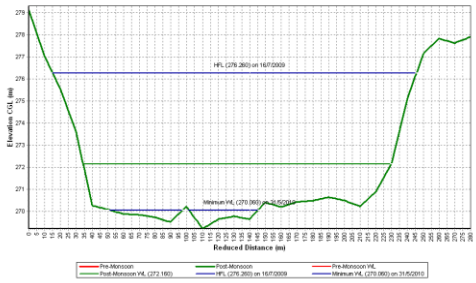
Ghatora



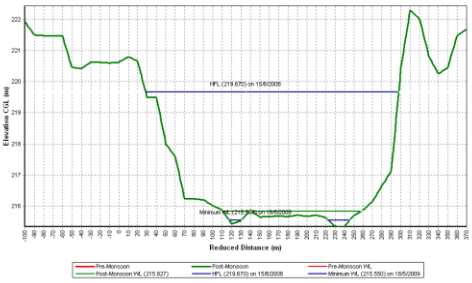
Jondhra



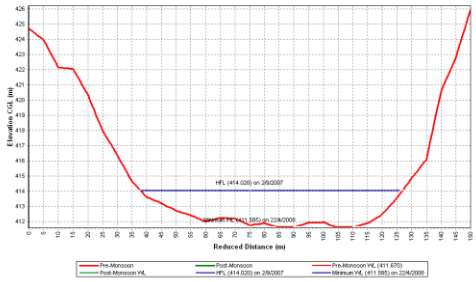
Kesinga



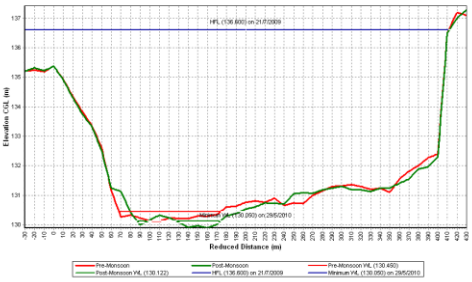
Kotni



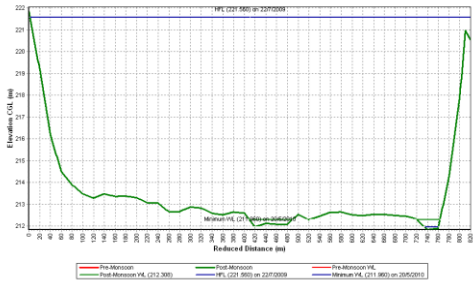
Kurubhata



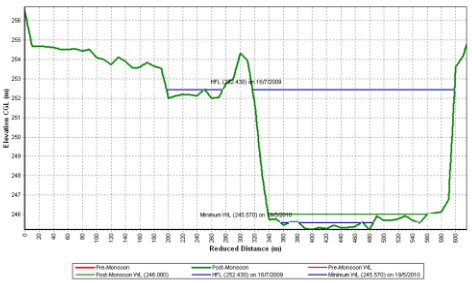
Manendragarh



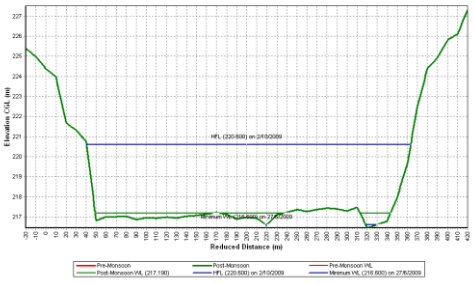
Salebhata



Seorinarayan

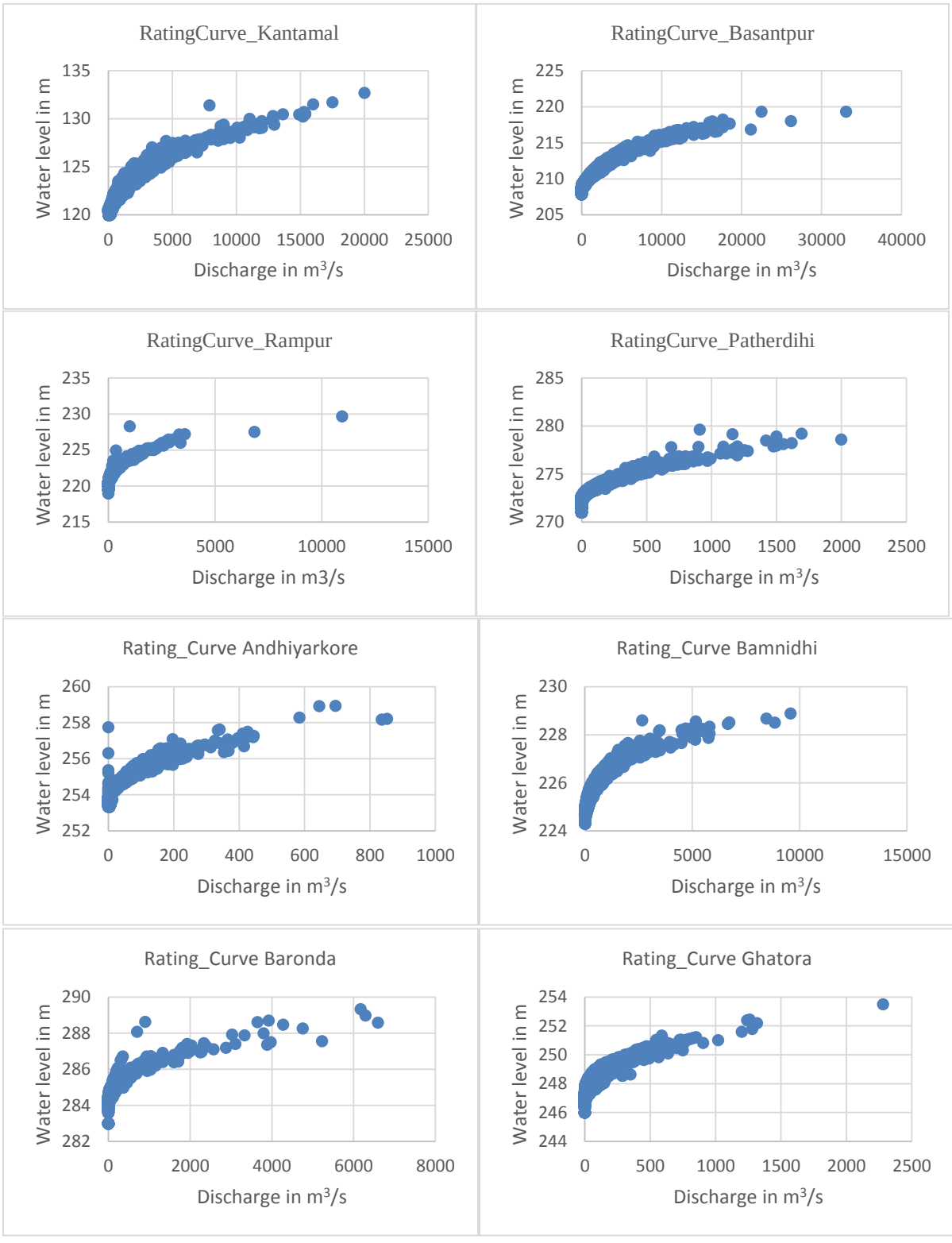


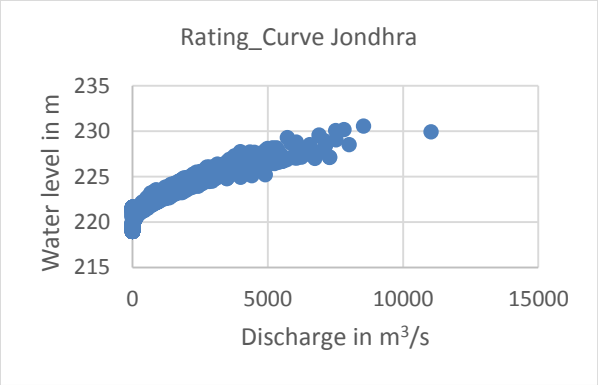
Simga

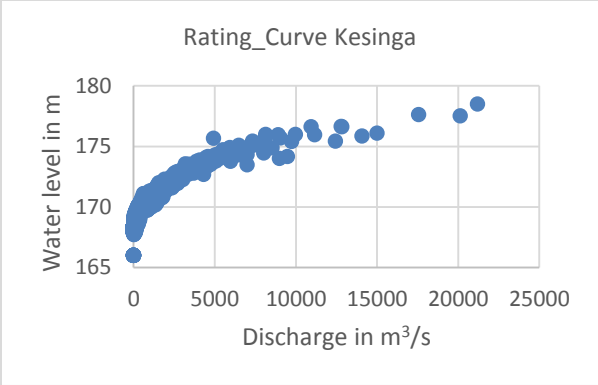


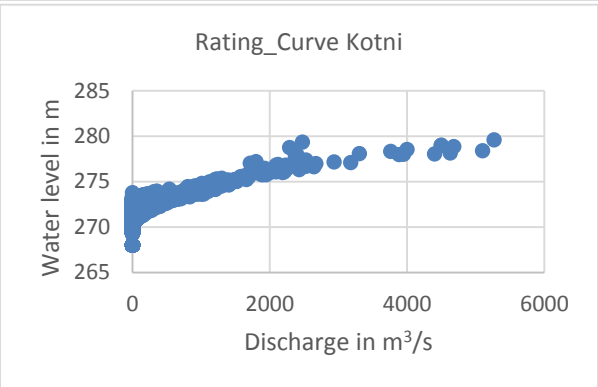
Sundargarh

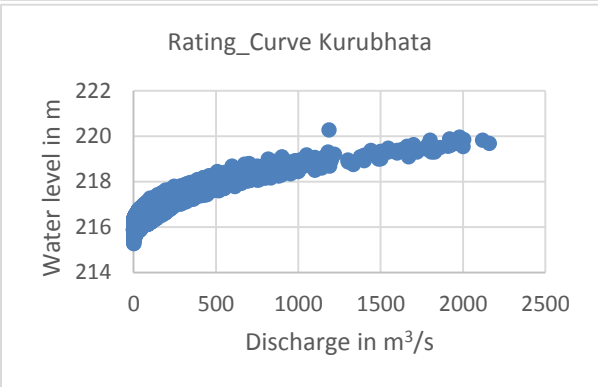
Cross-Section at four sampling locations of Mahanadi river system

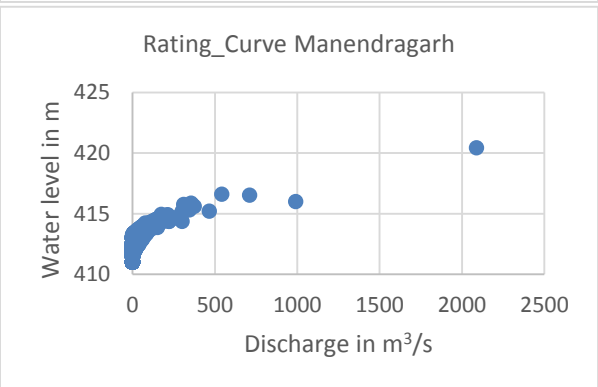

RatingCurve_Kantamal
Water level in m
135
130
125
120
0
5000
10000
15000
20000
25000
Discharge in m³/s
RatingCurve_Basantpur
Water level in m
225
220
215
210
205
0
10000
20000
30000
40000
Discharge in m³/s
RatingCurve_Rampur
Water level in m
235
230
225
220
215
0
5000
10000
15000
Discharge in m3/s
RatingCurve_Patherdihi
Water level in m
285
280
275
270
0
500
1000
1500
2000
2500
Discharge in m³/s
Rating_Curve Andhiyarkore
Water level in m
260
258
256
254
252
0
200
400
600
800
1000
Discharge in m³/s
Rating_Curve Bamnidhi
Water level in m
230
228
226
224
0
5000
10000
15000
Discharge in m³/s
Rating_Curve Baronda
Water level in m
290
288
286
284
282
0
2000
4000
6000
8000
Discharge in m³/s
Rating_Curve Ghatora
Water level in m
254
252
250
248
246
244
0
500
1000
1500
2000
2500
Discharge in m³/s

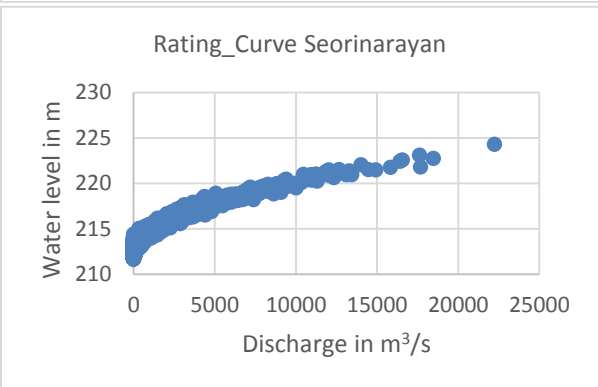

Rating_Curve Jondhra
Water level in m
235
230
225
220
215
0
5000
10000
15000
Discharge in m³/s

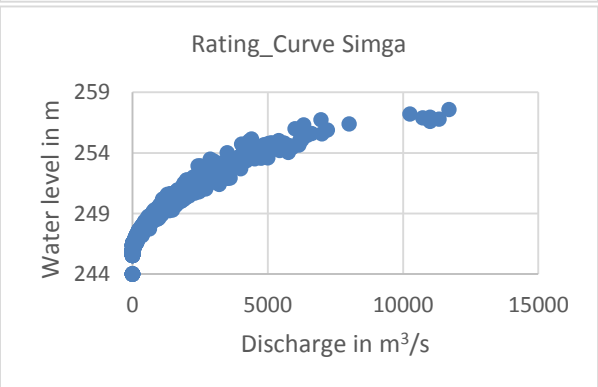

Rating_Curve Kesinga
Water level in m
180
175
170
165
0
5000
10000
15000
20000
25000
Discharge in m³/s

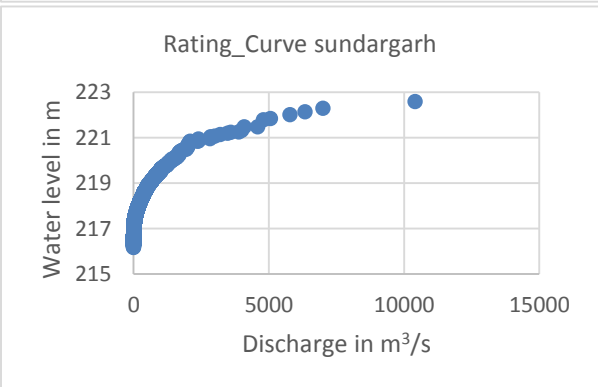

Rating_Curve Kotni
Water level in m
285
280
275
270
265
0
2000
4000
6000
Discharge in m³/s

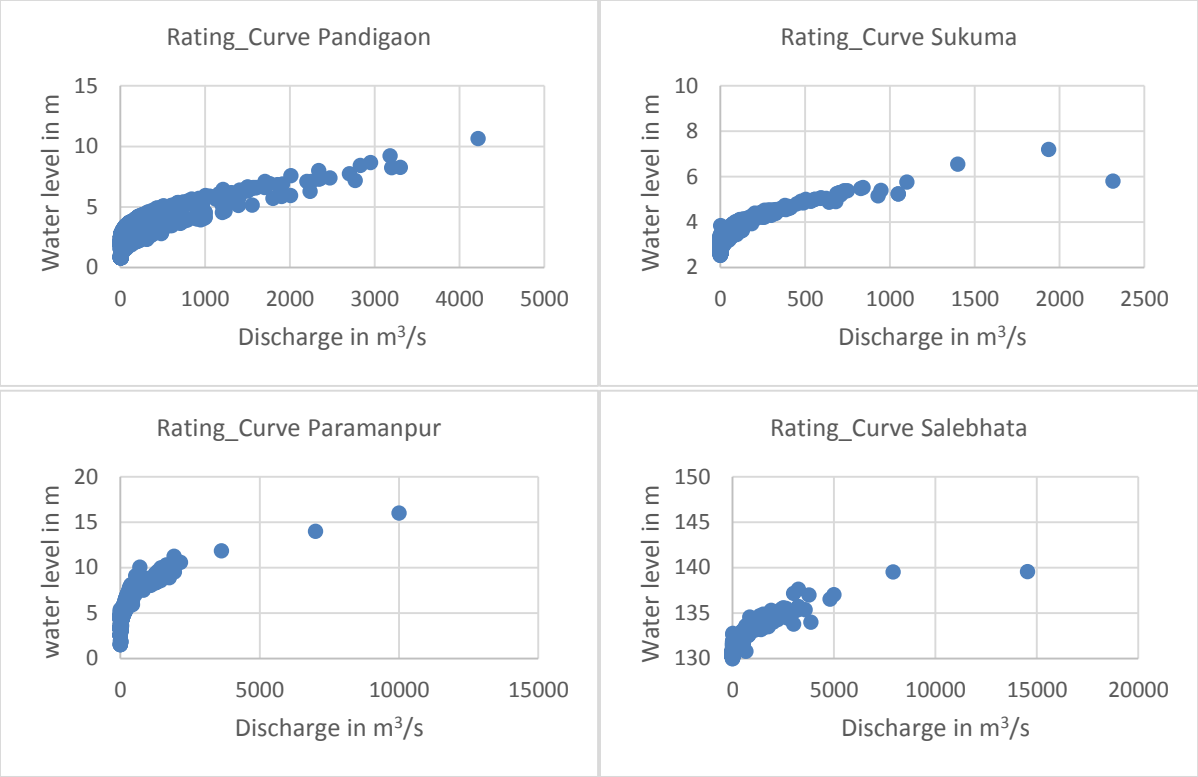

Rating_Curve Kurubhata
Water level in m
222
220
218
216
214
0
500
1000
1500
2000
2500
Discharge in m³/s


Rating_Curve Manendragarh
Water level in m
425
420
415
410
0
500
1000
1500
2000
2500
Discharge in m³/s


Rating_Curve Seorinarayan
Water level in m
230
225
220
215
210
0
5000
10000
15000
20000
25000
Discharge in m³/s


Rating_Curve Simga
Water level in m
259
254
249
244
0
5000
10000
15000
Discharge in m³/s


Rating_Curve sundargarh
Water level in m
223
221
219
217
215
0
5000
10000
15000
Discharge in m³/s

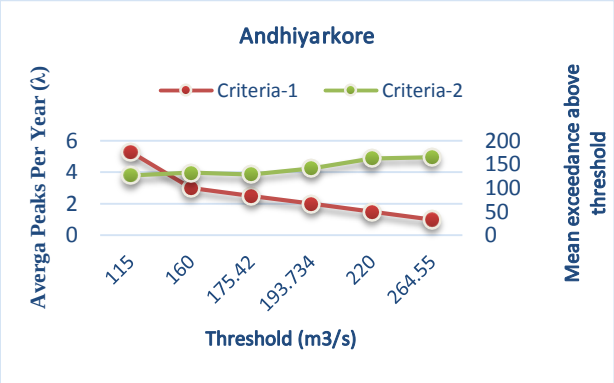


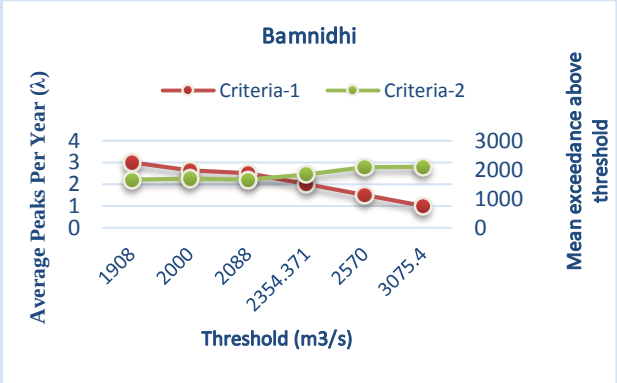
Rating curve at four sampling locations of Mahanadi river system

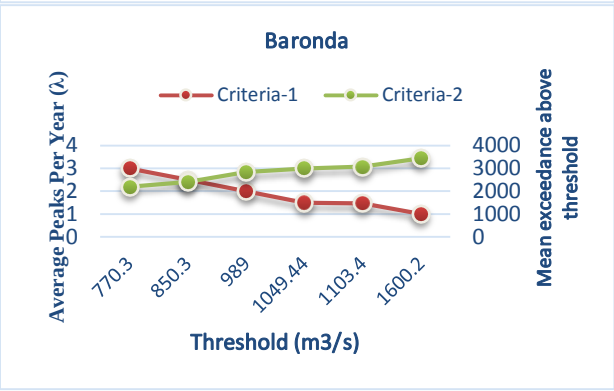
List of HFL value

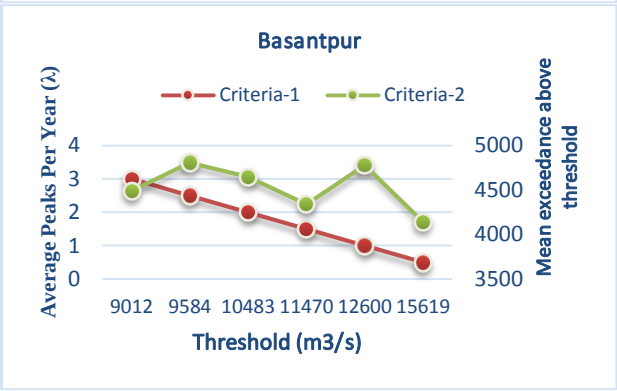
Station	HFL Value in m
Andhiyarkore	255.56
Bamnidhi	225.98
Baronda	286.16
Basantpur	216.89
Ghatora	247.90
Jhondhra	226.15
Kantamal	126.59
Kesinga	171.52
Kotni	276.23
Kurubhata	217.79
Manendragarh	414.03
Pandigaon	4.41
Paramanpur	6.2
Pathardihi	276.69
Rampur	225.01
Salebhata	136.55
Seorinarayan	221.17
Simga	252.48
Sukuma	5

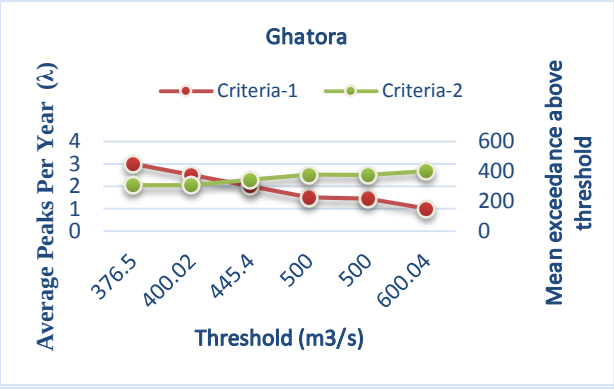
Sundargarh	220.64
Tikarapara	68.86
Rajim	279.72

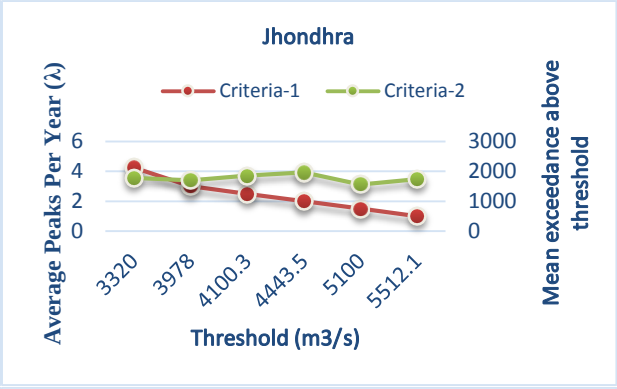


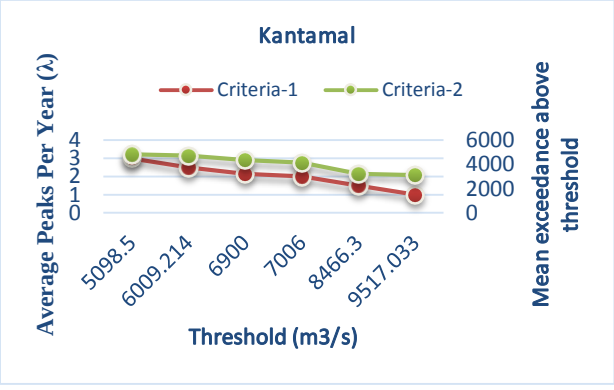


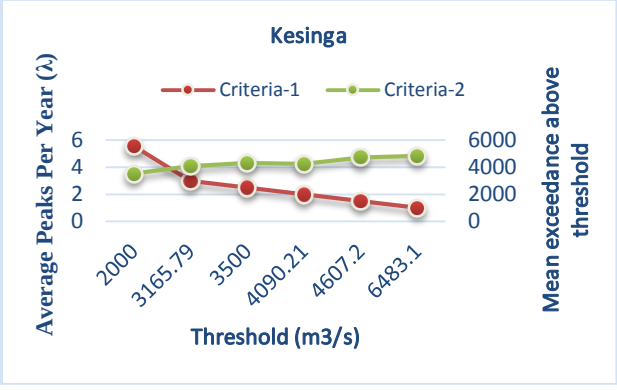


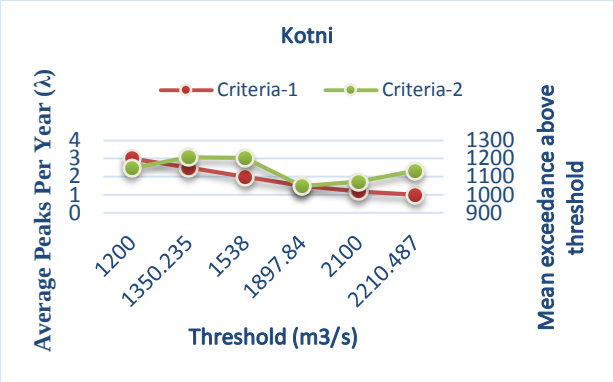


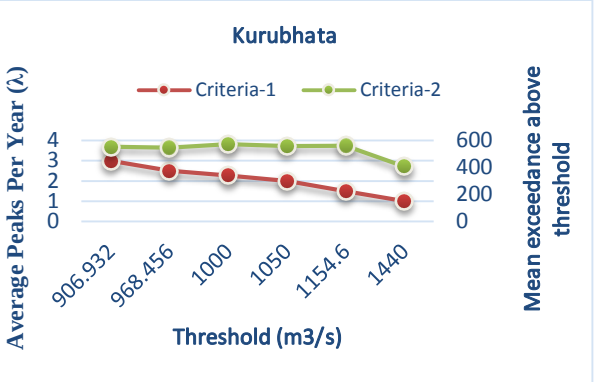


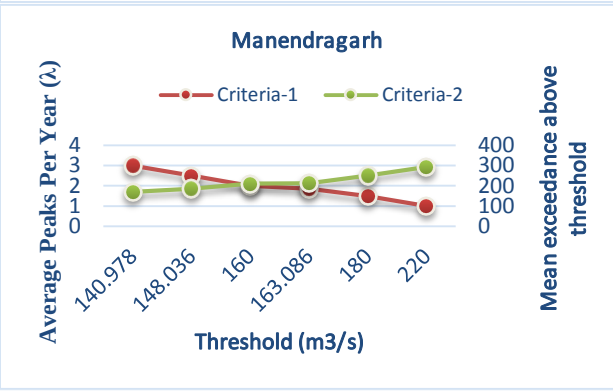


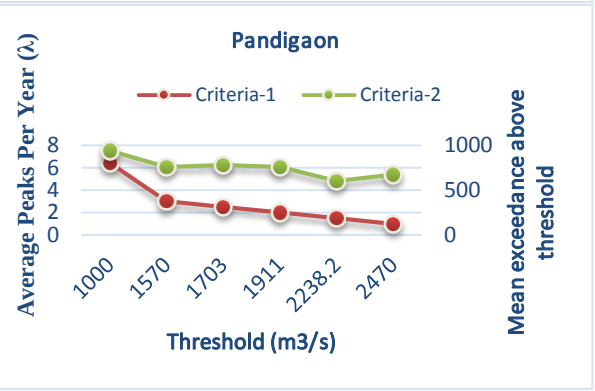


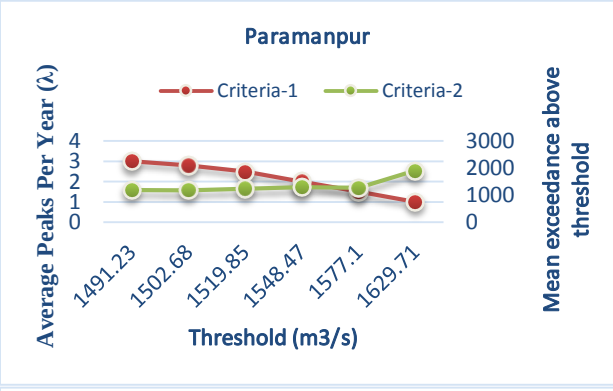


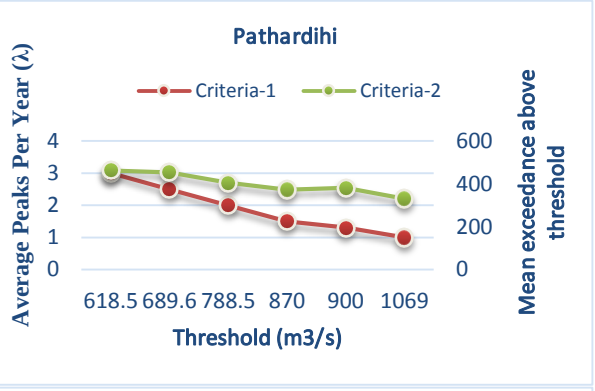

Kotni
Criteria-1
Criteria-2
Average Peaks Per Year (λ)
Mean exceedance above threshold
Threshold (m3/s)
4
3
2
1
0
1300
1200
1100
1000
900
1200
1350.235
1538
1897.84
2100
2210.487

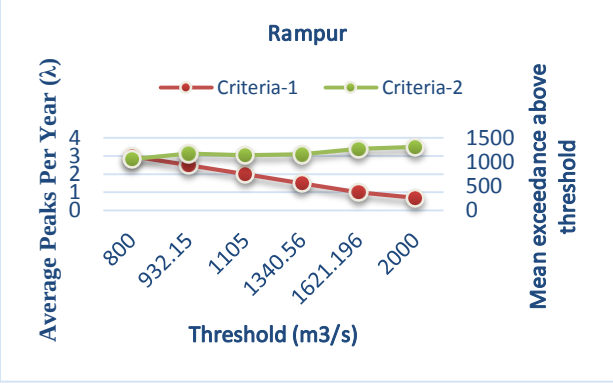

Kurubhata
Criteria-1
Criteria-2
Average Peaks Per Year (λ)
Mean exceedance above threshold
Threshold (m3/s)
4
3
2
1
0
600
400
200
0
906.932
968.456
1000
1050
1154.6
1440

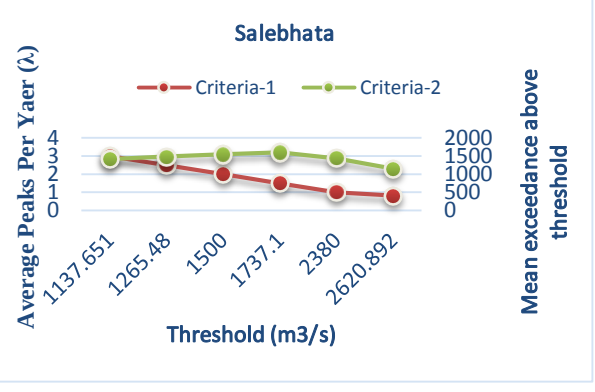

Manendragarh
Criteria-1
Criteria-2
Average Peaks Per Year (λ)
Mean exceedance above threshold
Threshold (m3/s)
4
3
2
1
0
400
300
200
100
0
140.978
148.036
160
163.086
180
220

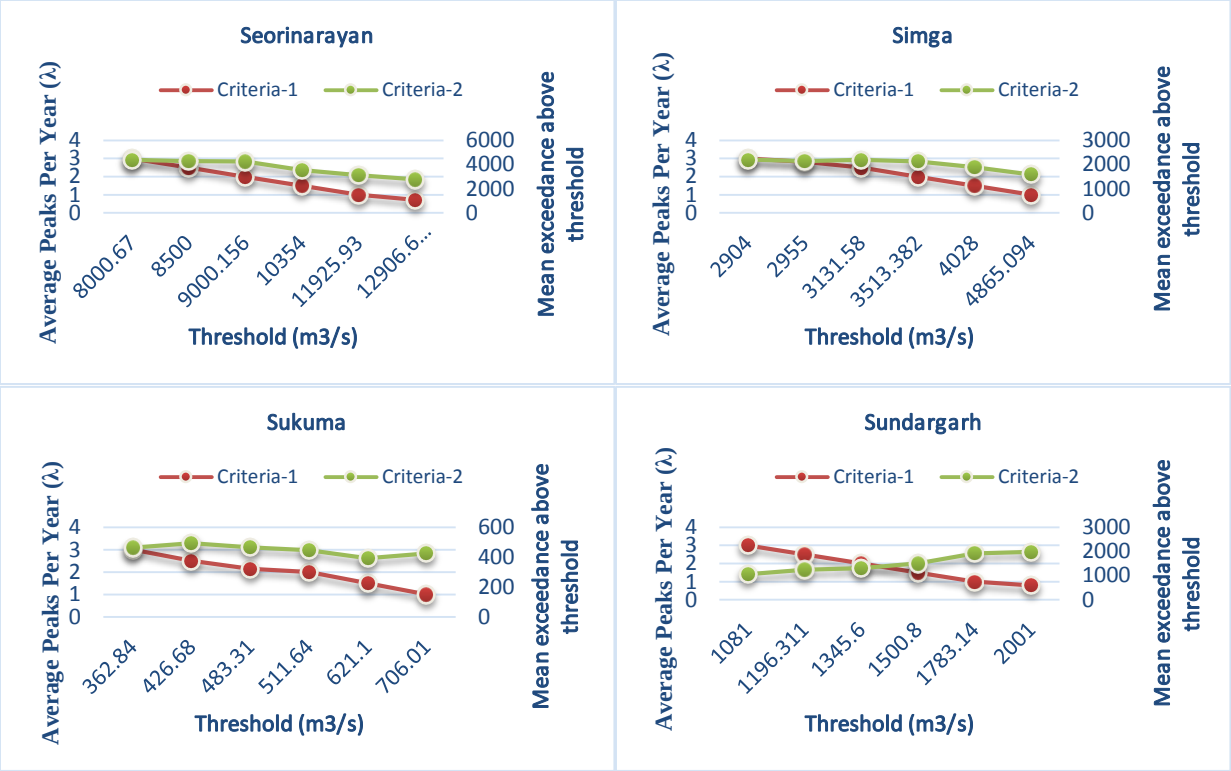

Pandigaon
Criteria-1
Criteria-2
Average Peaks Per Year (λ)
Mean exceedance above threshold
Threshold (m3/s)
8
6
4
2
0
1000
500
0
1000
1570
1703
1911
2238.2
2470


Paramanpur
Criteria-1
Criteria-2
Average Peaks Per Year (λ)
Mean exceedance above threshold
Threshold (m3/s)
4
3
2
1
0
3000
2000
1000
0
1491.23
1502.68
1519.85
1548.47
1577.1
1629.71

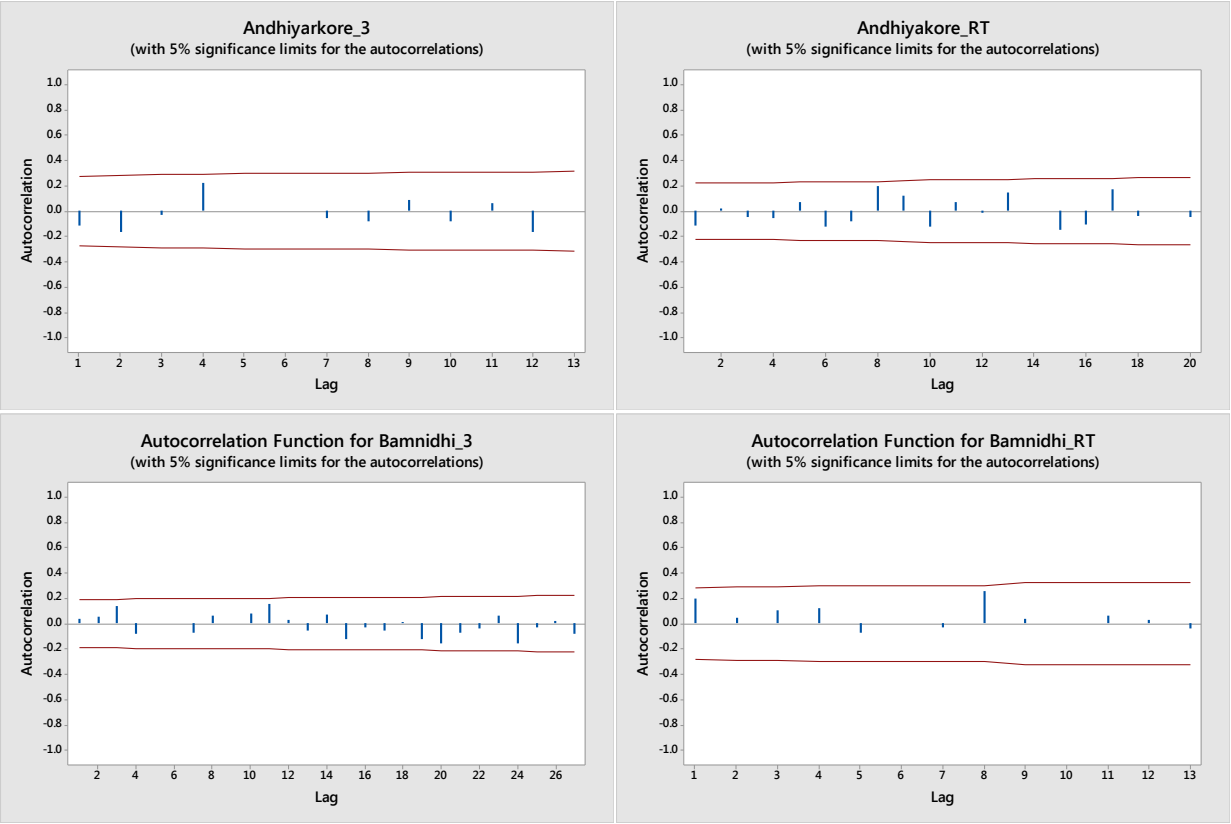

Pathardihi
Criteria-1
Criteria-2
Average Peaks Per Year (λ)
Mean exceedance above threshold
Threshold (m3/s)
4
3
2
1
0
600
400
200
0
618.5
689.6
788.5
870
900
1069


Rampur
Criteria-1
Criteria-2
Average Peaks Per Year (λ)
Mean exceedance above threshold
Threshold (m3/s)
4
3
2
1
0
1500
1000
500
0
800
932.15
1105
1340.56
1621.196
2000

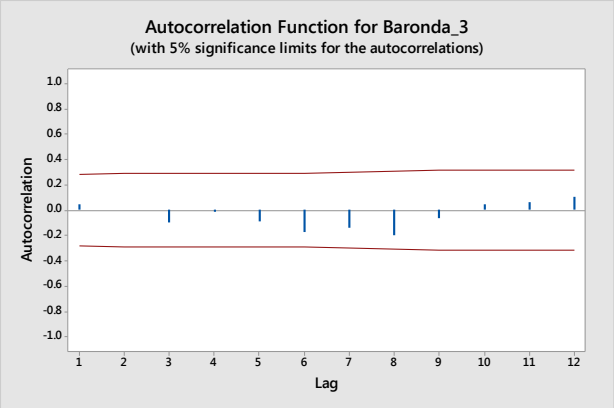

Salebhata
Criteria-1
Criteria-2
Average Peaks Per Yaer (λ)
Mean exceedance above threshold
Threshold (m3/s)
4
3
2
1
0
2000
1500
1000
500
0
1137.651
1265.48
1500
1737.1
2380
2620.892

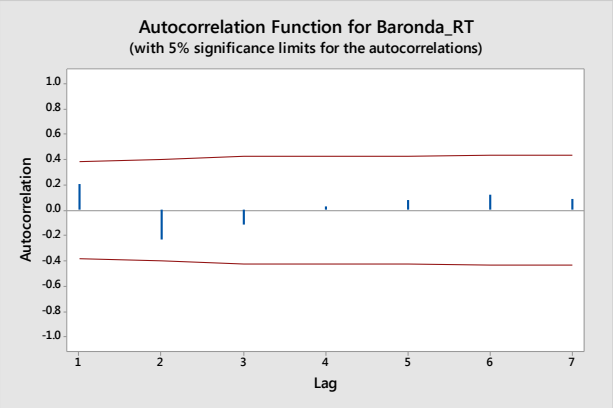


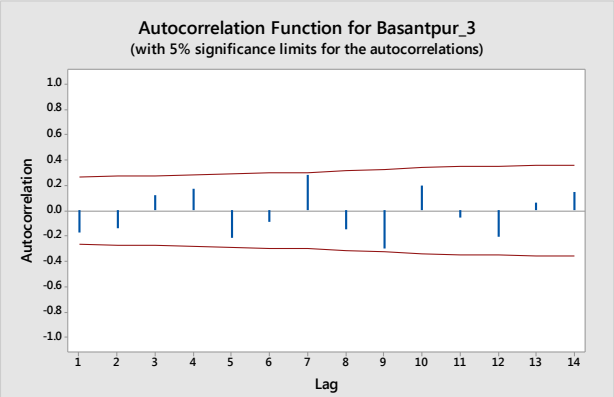
Threshold values obtained using different concepts

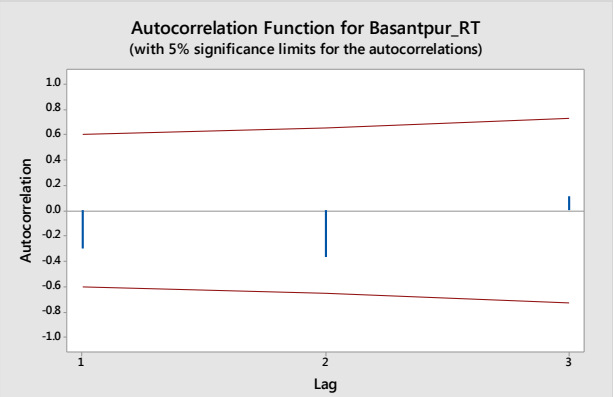


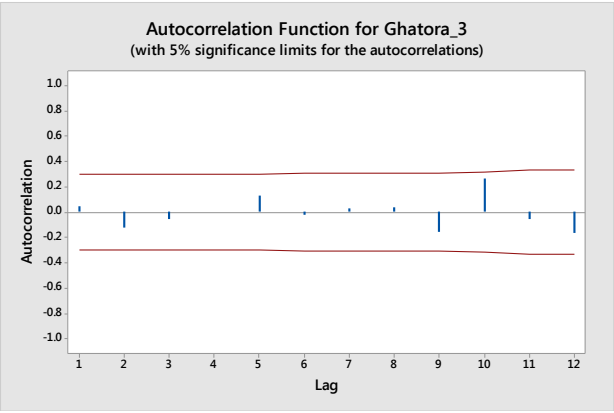
X

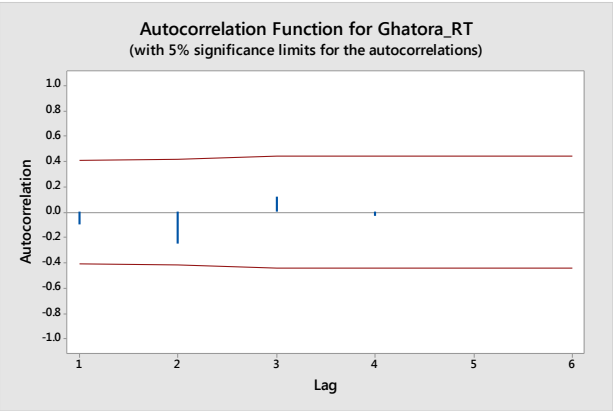

Autocorrelation Function for Baronda_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

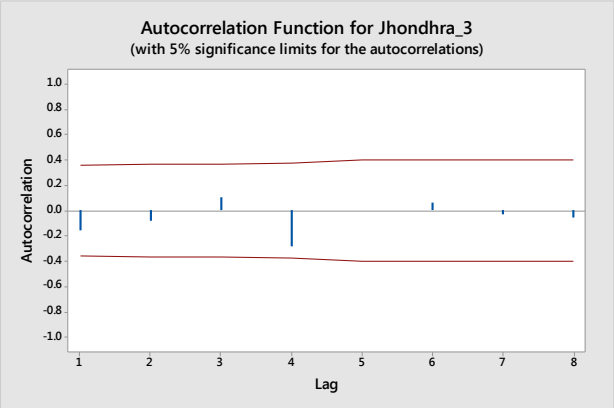

Autocorrelation Function for Baronda_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

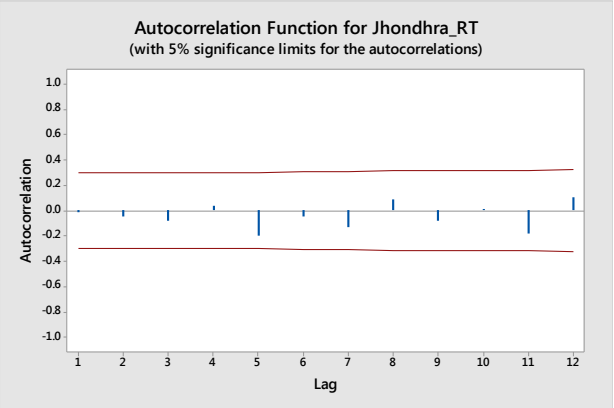

Autocorrelation Function for Basantpur_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

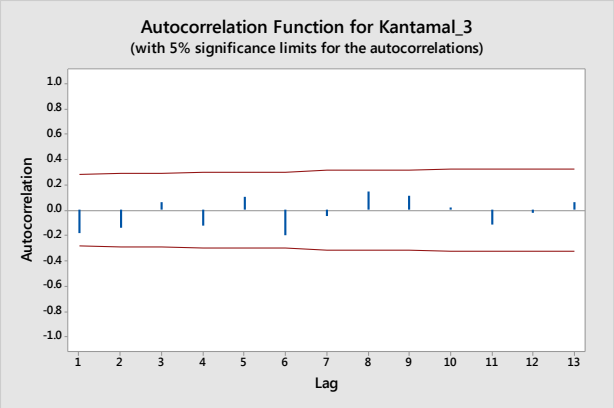

Autocorrelation Function for Basantpur_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

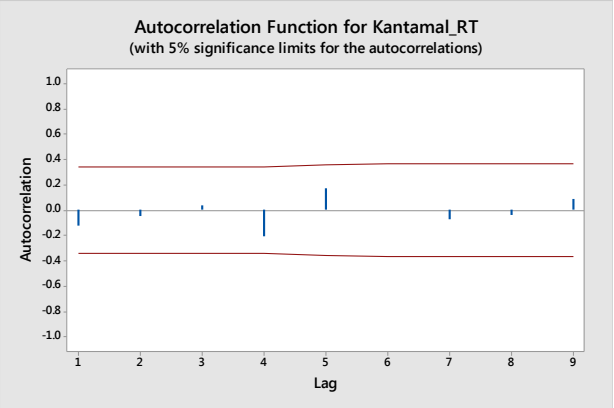

Autocorrelation Function for Ghatora_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

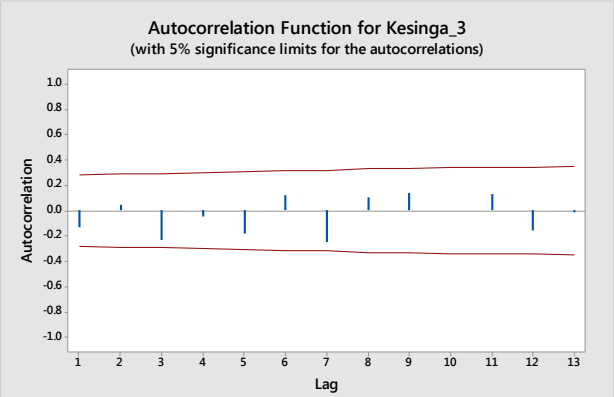

Autocorrelation Function for Ghatora_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

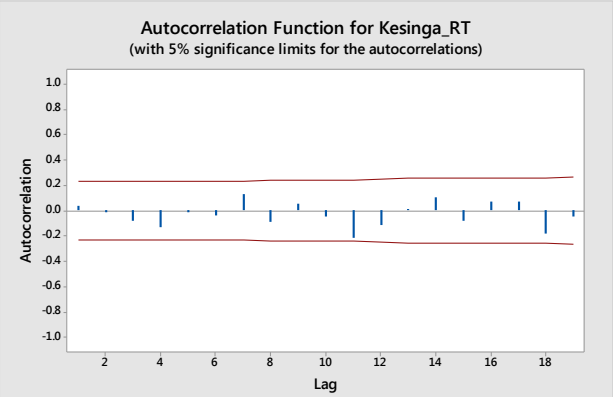

Autocorrelation Function for Jhondhra_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

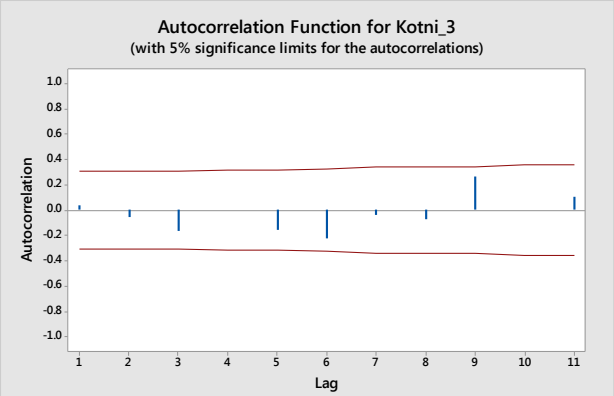

Autocorrelation Function for Jhondhra_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

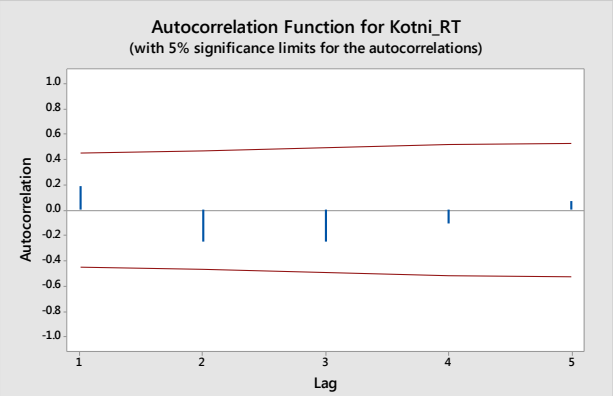

Autocorrelation Function for Kantamal_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

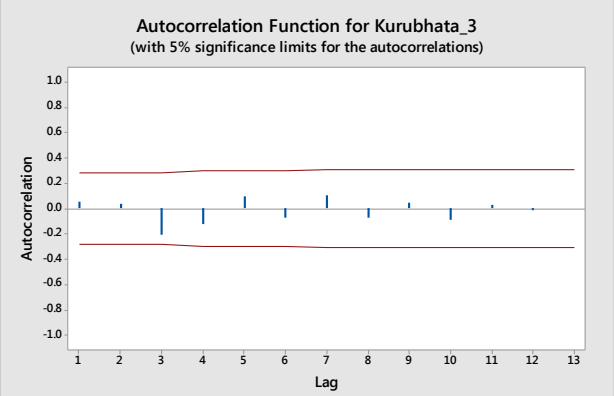

Autocorrelation Function for Kantamal_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

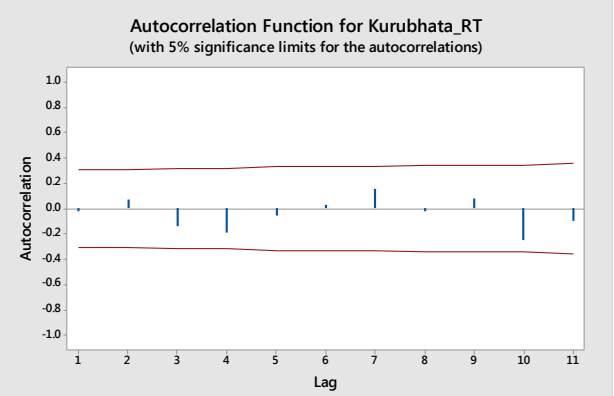

Autocorrelation Function for Kesinga_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

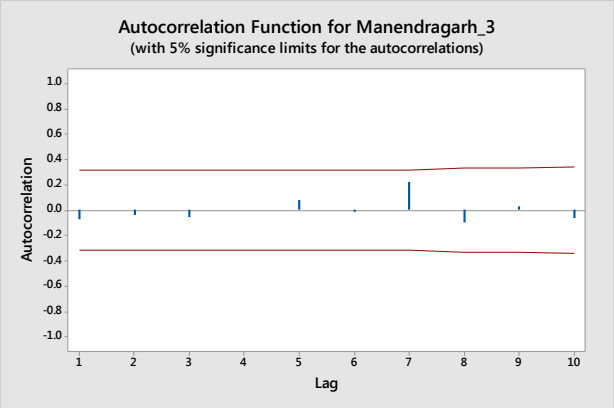

Autocorrelation Function for Kesinga_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

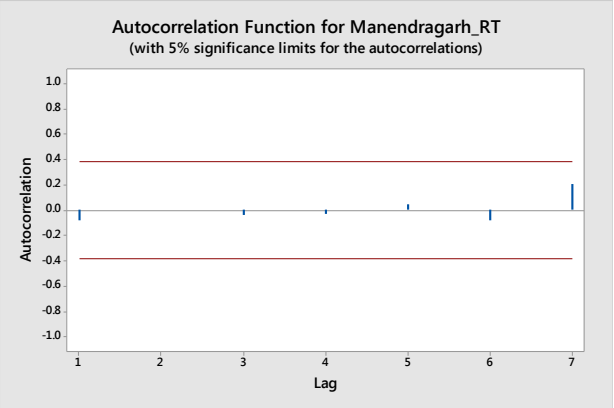

Autocorrelation Function for Kotni_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

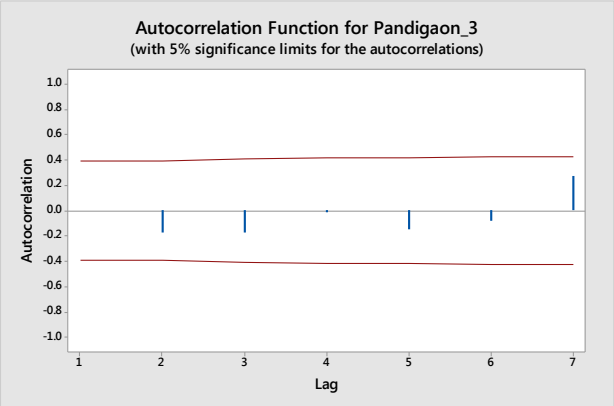

Autocorrelation Function for Kotni_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

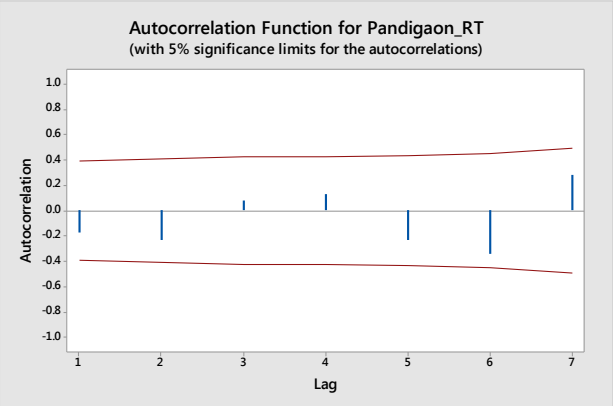

Autocorrelation Function for Kurubhata_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

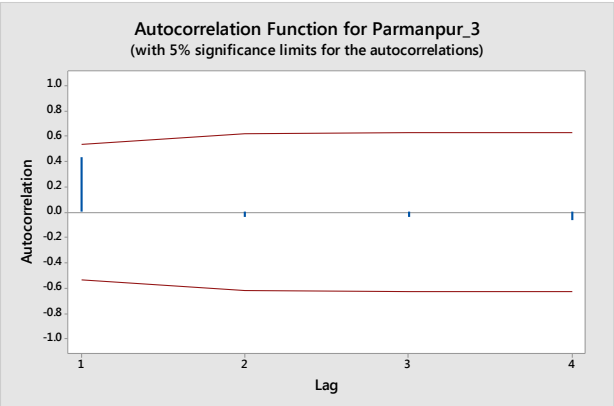

Autocorrelation Function for Kurubhata_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

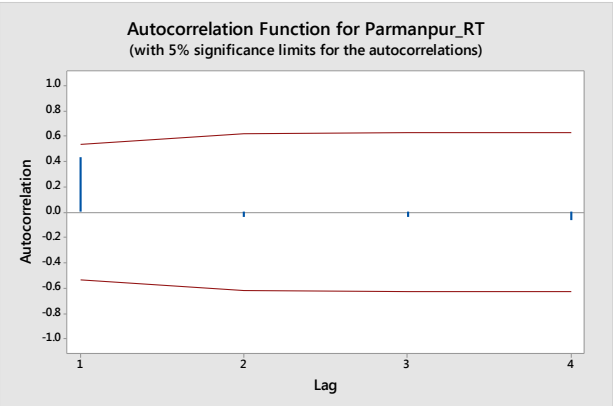

Autocorrelation Function for Manendragarh_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

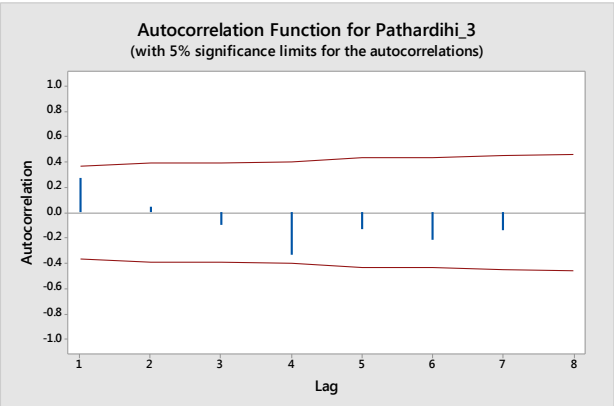

Autocorrelation Function for Manendragarh_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

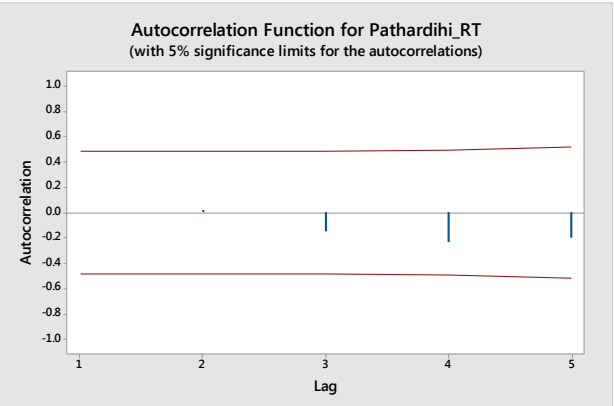

Autocorrelation Function for Pandigaon_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

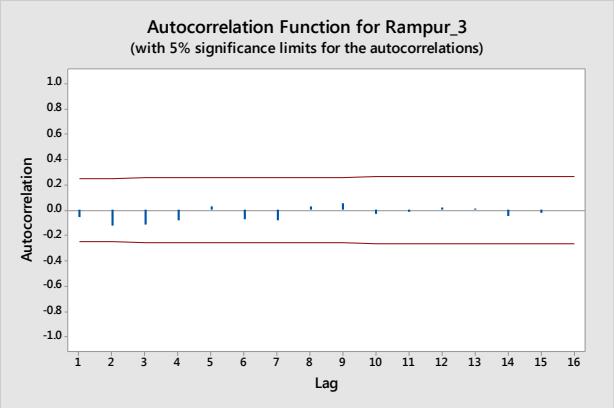

Autocorrelation Function for Pandigaon_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

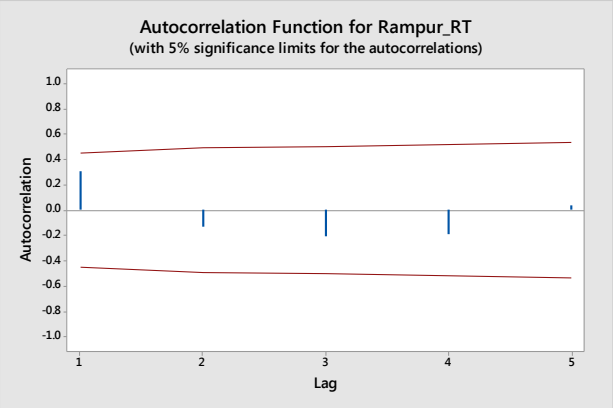

Autocorrelation Function for Parmanpur_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

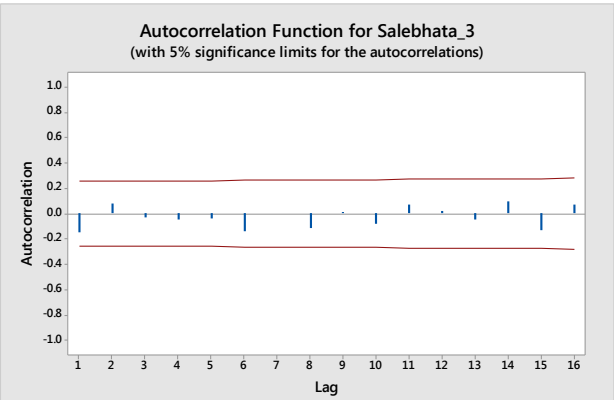

Autocorrelation Function for Parmanpur_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

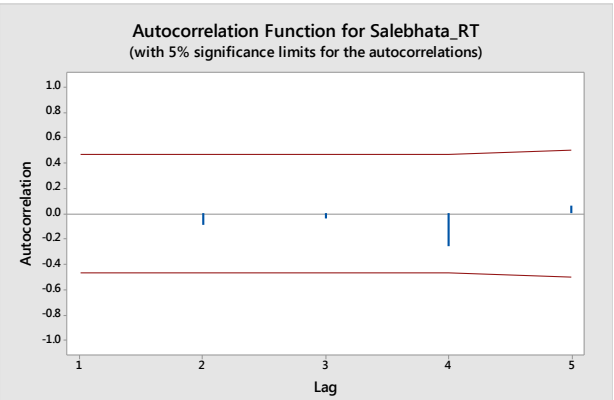

Autocorrelation Function for Pathardihi_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

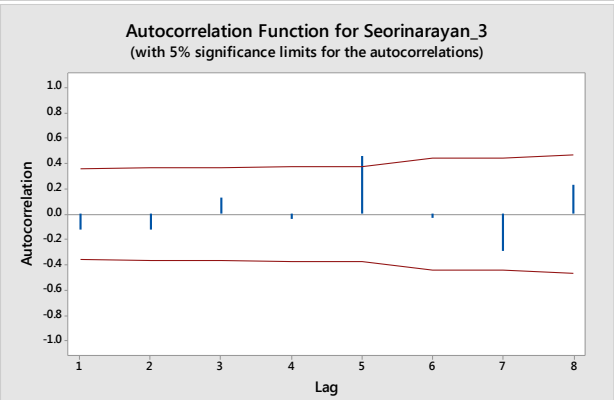

Autocorrelation Function for Pathardihi_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

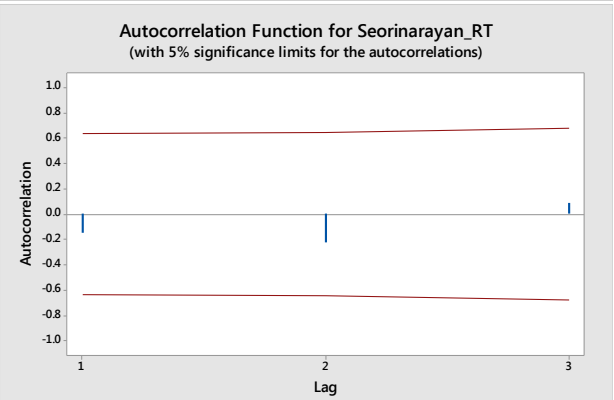

Autocorrelation Function for Rampur_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

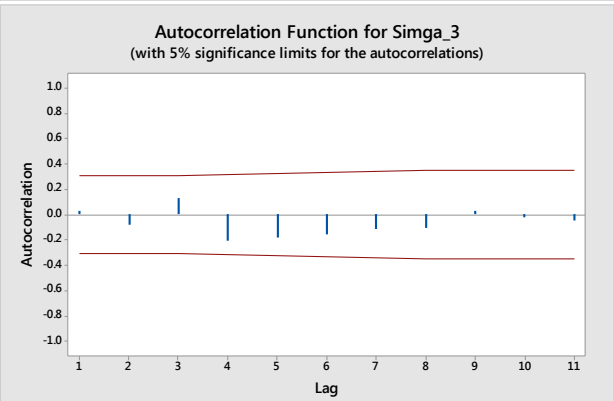

Autocorrelation Function for Rampur_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

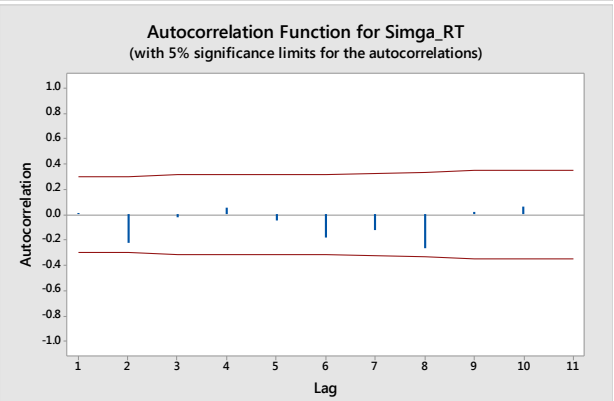

Autocorrelation Function for Salebhata_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag

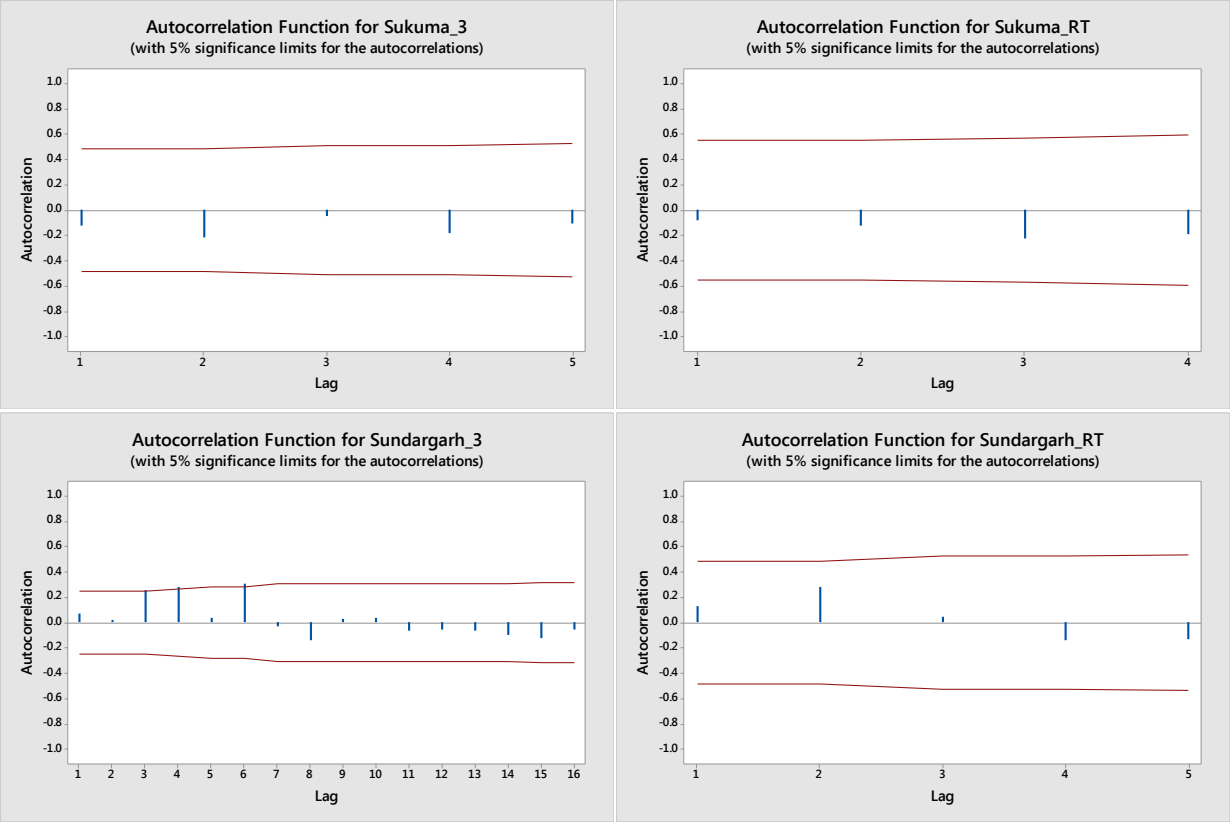

Autocorrelation Function for Salebhata_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag


Autocorrelation Function for Seorinarayan_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag


Autocorrelation Function for Seorinarayan_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag


Autocorrelation Function for Simga_3
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag


Autocorrelation Function for Simga_RT
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag



Autocorrelation function of the discharge data

Kendell Trend Test

Station	Kendell Trend Test (Z-Statistics)					
	1	1.5	2	2.5	3	HFL
Andhiyarkore	0.395 (NO)	0.296 (NO)	0.327 (NO)	0.147 (NO)	0.971 (NO)	0.952 (NO)
Bamnidhi	0.298 (NO)	-0.867 (NO)	-1.182 (NO)	-1.363 (NO)	-1.353 (NO)	-1.564 (NO)
Baronda	1.241 (NO)	0.625 (NO)	1.596 (NO)	-0.065 (NO)	0.155 (NO)	0.257 (NO)
Basantpur	0.163 (NO)	-0.586 (NO)	-1.387 (NO)	0.19 (NO)	0.179 (NO)	-1.281 (NO)
Ghatora	-1.061 (NO)	0.49 (NO)	0 (NO)	-0.034 (NO)	-0.436 (NO)	0.444 (NO)
Jhondhra	-0.985 (NO)	-0.552 (NO)	-1.056 (NO)	-1.182 (NO)	-0.049 (NO)	0.663 (NO)
Kantamal	0.911 (NO)	-0.17 (NO)	0.83 (NO)	1.088 (NO)	-0.049 (NO)	0.341 (NO)

Kesinga	0.097 (NO)	0.281 (NO)	-0.85 (NO)	-0.82 (NO)	0.335 (NO)	0.067 (NO)
Kotni	-1.591 (Decreasing)	-0.882 (NO)	-1.156 (NO)	-0.341 (NO)	0.586 (NO)	-0.634 (NO)
Kurubhata	-0.264 (NO)	0.325 (NO)	0.236 (NO)	0.596 (NO)	-0.293 (NO)	0.022 (NO)
Manendragarh	-0.453 (NO)	-1.637 (NO)	-2.319 (Decreasing)	-2.452 (Decreasing)	-1.898 (Decreasing)	-2.045 (Decreasing)
Pandigaon	0.001 (NO)	1.037 (NO)	0.783 (NO)	1.41 (NO)	0.5 (NO)	0.479 (NO)
Paramanpur	-1.147 (Decreasing)	0.0001 (NO)	0.0011 (NO)	0.297 (NO)	0.585 (NO)	0.585 (NO)
Pathardihi	0.602 (NO)	-0.302 (NO)	-0.645 (NO)	-1.313 (NO)	-1.003 (NO)	0.385 (NO)
Rampur	0.506 (NO)	1.13 (NO)	1.623 (NO)	1.421 (NO)	0.504 (NO)	0.453 (NO)
Salebhata	0.896 (NO)	0.811 (NO)	1.074 (NO)	1.067 (NO)	1.148 (NO)	1.038 (NO)
Seorinarayan	-0.792 (NO)	1.297 (NO)	1.252 (NO)	1.245 (NO)	1.087 (NO)	-0.617 (NO)
Simga	0.158 (NO)	0.428 (NO)	-1.193 (NO)	-0.641 (NO)	-0.858 (NO)	-1.153 (NO)
Sukuma	0.069 (NO)	1.037 (NO)	0.396 (NO)	0.045 (NO)	0.14 (NO)	0.594 (NO)
Sundargarh	0.151 (NO)	1.362 (NO)	0.878 (NO)	1.297 (NO)	0.267 (NO)	1.469 (NO)
Tikarapara	0.242 (NO)	-0.506 (NO)	0.341 (NO)	0.011 (NO)	-0.663 (NO)	-1.281 (NO)
Rajim	-1.363 (NO)	-1.244 (NO)	-1.376 (NO)	-1.062 (NO)	-1.032 (NO)	-1.129 (NO)

Shape Parameter for the PDS peak floods

Station	**Average Peaks per Years (λ) Shape Parameter**				
	1	1.5	2	2.5	3
Andhiyarkore	-0.182	-0.206	-0.281	-0.337	-0.2639
Bamnidhi	-0.044	-0.010	-0.084	-0.1829	-0.186
Baronda	0.345	0.089	0.184	-0.058	-0.163

Basantpur	-0.203	-0.135	-0.082	-0.008	-0.004
Ghatora	-0.286	-0.309	-0.297	-0.320	-0.303
Jhondhra	-0.331	-0.287	-0.092	-0.087	0.040
Kantamal	0.103	0.212	0.336	0.474	0.529
Kesinga	0.102	0.107	-0.039	-0.103	-0.091
Kotni	0.274	0.014	0.021	0.031	0.032
Kurubhata	1.056	1.669	1.177	0.826	0.670
Manendragarh	-0.609	-0.560	-0.558	-0.564	-0.578
Pandigaon	-0.127	0.165	0.185	0.240	0.215
Paramanpur	-0.556	-0.695	-0.693	-0.707	-0.727
Pathardihi	0.358	0.371	0.268	0.299	0.33
Rampur	-0.378	-0.319	-0.294	-0.239	-0.145
Salebhata	-0.445	-0.328	-0.259	-0.209	-0.224
Seorinarayan	-0.167	0.066	0.249	0.297	0.422
Simga	-0.239	-0.133	-0.029	0.095	0.132
Sukuma	-0.242	-0.462	-0.454	-0.3386	-0.223
Sundargarh	-0.069	-0.235	-0.274	-0.344	-0.404
Tikarapara	1.061	0.895	0.786	0.696	0.608
Rajim	0.805	0.451	0.369	0.286	0.236

Scale Parameter for the PDS peak floods

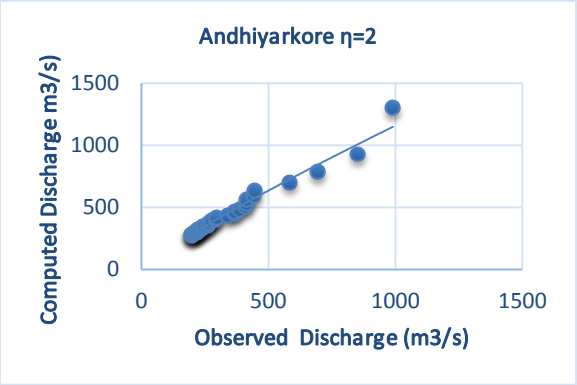
Station	**Average Peaks per Years (λ) Scale Parameter**				
	1	1.5	2	2.5	3
Andhiyarkore	140.864	131.924	105.1744	89.295	101.006
Bamnidhi	1904.397	2010.959	1769.048	1350.347	1341.402
Baronda	2880.310	2089.675	2332.334	1566.138	1250.224
Basantpur	3291.406	3597.397	3929.068	4467.577	4658.218
Ghatora	299.313	210.259	245.531	215.199	215.116
Jhondhra	1124.830	1124.607	1639.600	1645.075	1807.885
Kantamal	3141.039	4022.218	5037.946	6227.057	7328.459

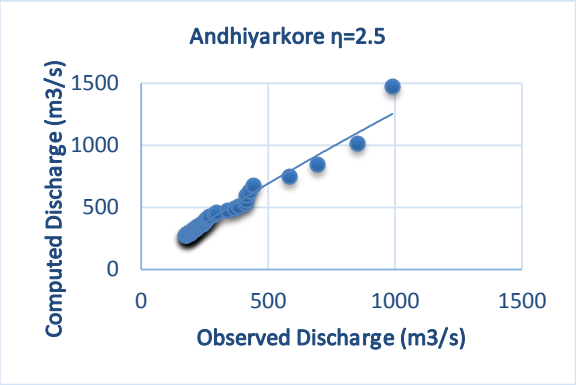
Kesinga	5563.978	5960.979	4543.972	4049.017	3967.483
Kotni	1710.012	1164.120	1187.885	1222.454	1233.264
Kurubhata	683.209	1725.281	1429.847	1176.190	1038.518
Manendragarh	108.334	108.214	93.967	83.250	73.721
Pandigaon	426.309	763.380	817.739	942.778	949.665
Paramanpur	941.7183	421.830	426.877	381.183	331.604
Pathardihi	483.834	580.741	520.072	563.213	619.472
Rampur	751.006	779.991	773.887	854.349	886.768
Salebhata	796.413	955.777	1058.357	1137.518	1066.136
Seorinarayan	2277.569	3403.543	4853.373	5378.550	6393.562
Simga	1244.973	1509.037	1794.034	2222.660	2440.208
Sukuma	341.050	214.977	206.493	275.0422	349.820
Sundargarh	1855.854	1238.374	1057.785	848.872	633.996
Tikarapara	10543.860	11612.342	12137.121	11816.196	11481.224
Rajim	5298.809	4015.724	3839.487	3403.519	3168.840

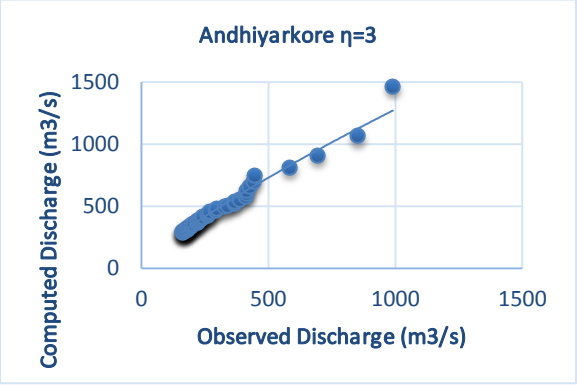
Different Parameters of PDS obtained from HFL Concept

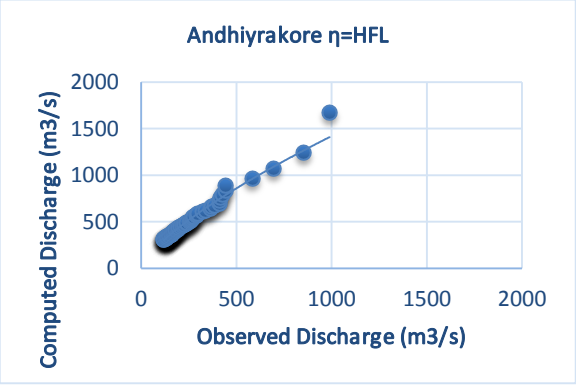
Station	**Threshold Value (t)**	**Average Peaks per Year (λ)**	**HFL**	
			Shape	Scale
Andhiyarkore	115	5.29	-0.243	94.698
Bamnidhi	2000	2.63	-0.180	1383.067
Baronda	1103.4	1.47	0.168	2264.201
Basantpur	15619	0.5	-0.307	3130.271
Ghatora	500	1.45	-0.311	255.643
Jhondhra	3320	4.25	0.058	1873.037
Kantamal	6900	2.15	0.181	4022.726
Kesinga	2000	5.57	-0.233	2741.784
Kotni	2100	1.18	0.137	1420.437
Kurubhata	1000	2.27	1.045	1336.142
Manendragarh	163.08	1.86	-0.588	63.145

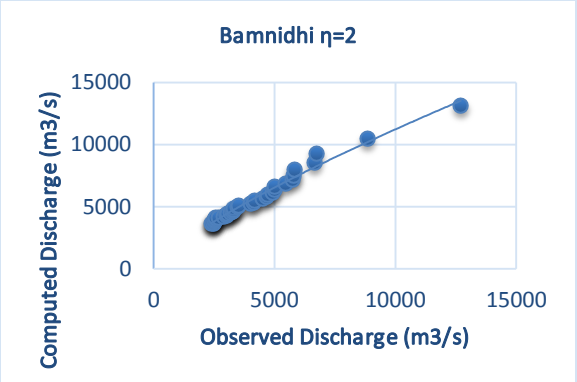
Pandigaon	1000	6.4	0.133	1011.699
Paramanpur	1502.68	2.8	-0.729	328.550
Pathardihi	900	1.3	0.427	604.124
Rampur	2000	0.7	-0.475	658.617
Salebhata	2620.892	0.82	-0.034	1093.348
Seorinarayan	12906.61	0.72	-0.283	1867.986
Simga	2955	2.82	0.082	2231.101
Sukuma	483.31	2.15	-0.375	256.462
Sundargarh	2001	0.79	-0.045	1886.524
Tikarapara	17741.25	2.85	0.562	10925.876
Rajim	2256.82	2.12	0.345	3729.371

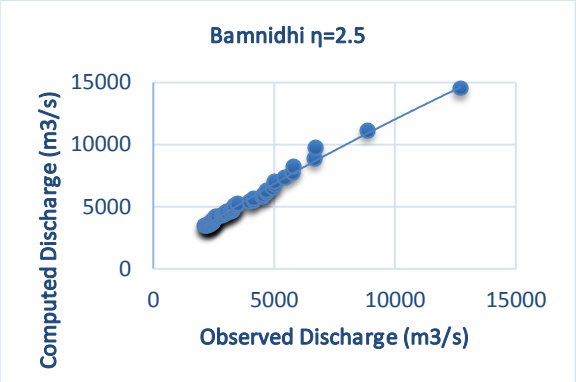


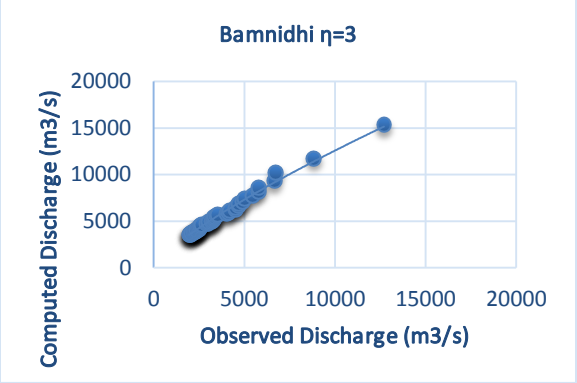


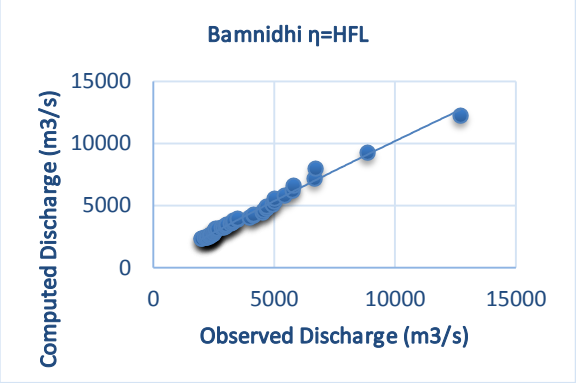


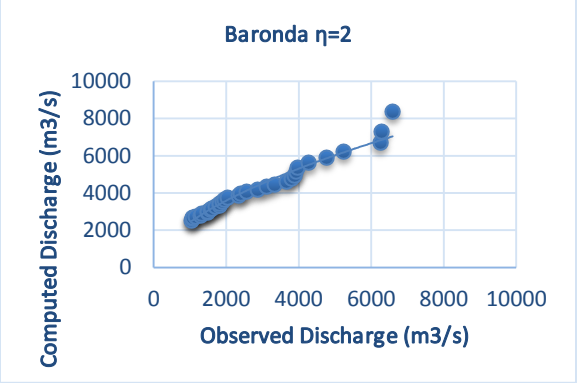


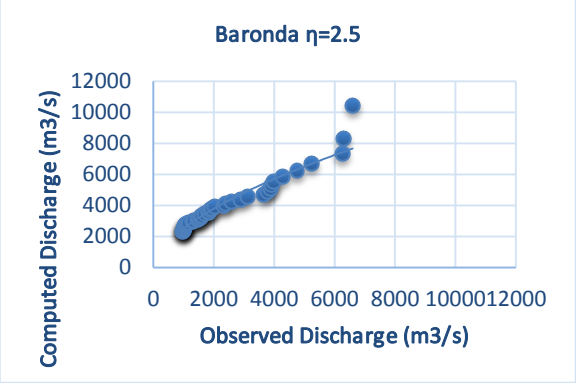

Bamnidhi η=2
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

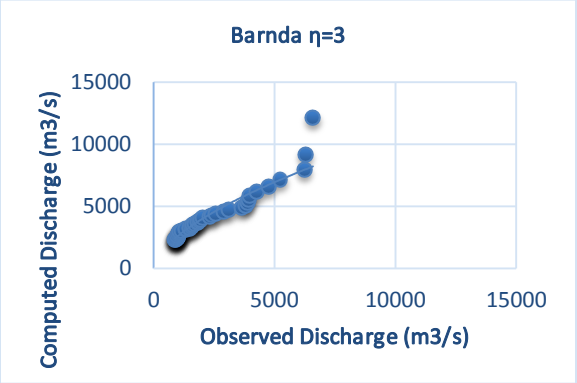

Bamnidhi η=2.5
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

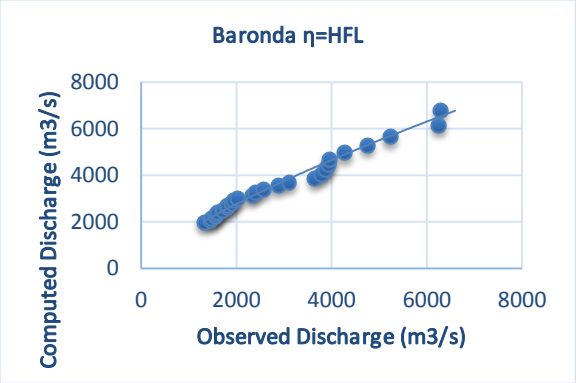

Bamnidhi η=3
Computed Discharge (m3/s)
20000
15000
10000
5000
0
0
5000
10000
15000
20000
Observed Discharge (m3/s)

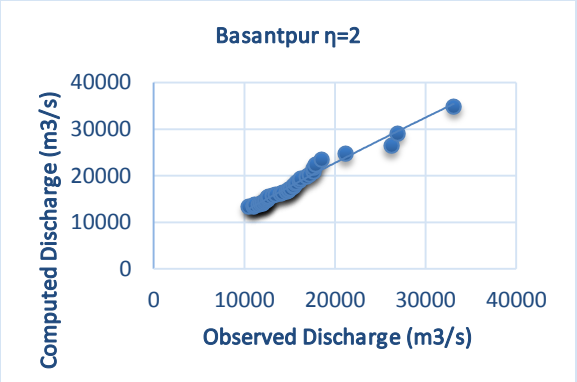

Bamnidhi η=HFL
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

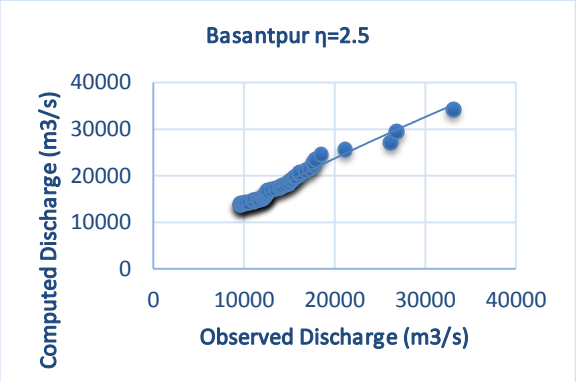

Baronda η=2
Computed Discharge (m3/s)
10000
8000
6000
4000
2000
0
0
2000
4000
6000
8000
10000
Observed Discharge (m3/s)

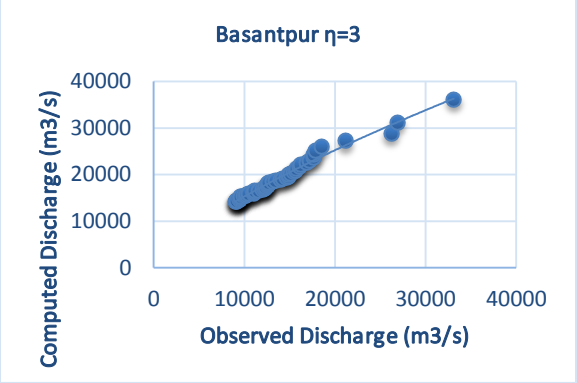

Baronda η=2.5
Computed Discharge (m3/s)
12000
10000
8000
6000
4000
2000
0
0
2000
4000
6000
8000
10000
12000
Observed Discharge (m3/s)

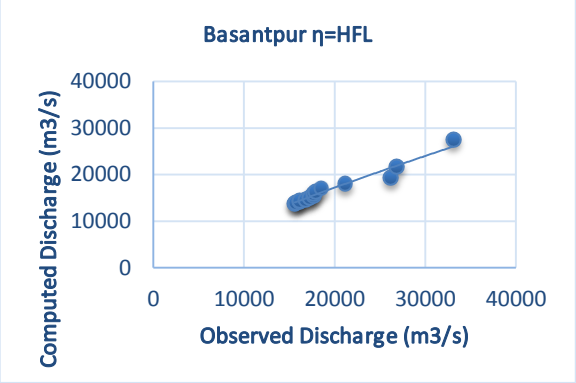

Barnda η=3
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

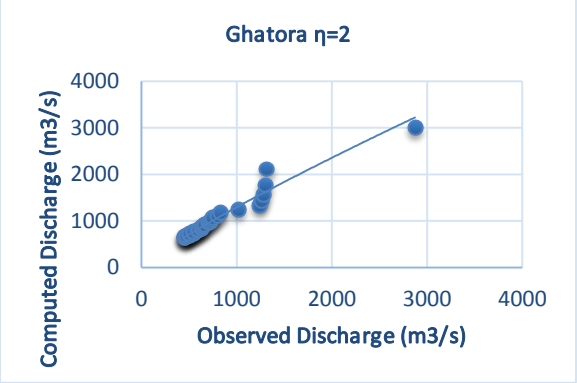

Baronda η=HFL
Computed Discharge (m3/s)
8000
6000
4000
2000
0
0
2000
4000
6000
8000
Observed Discharge (m3/s)

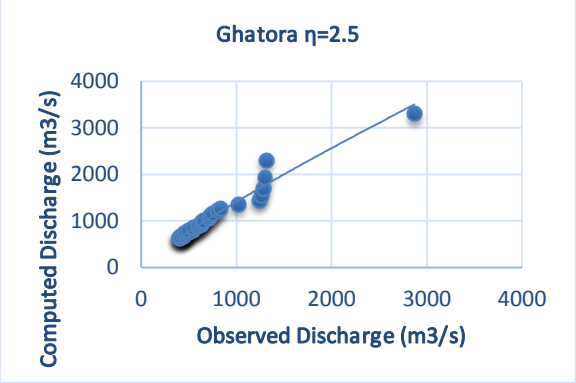

Basantpur η=2
Computed Discharge (m3/s)
40000
30000
20000
10000
0
0
10000
20000
30000
40000
Observed Discharge (m3/s)

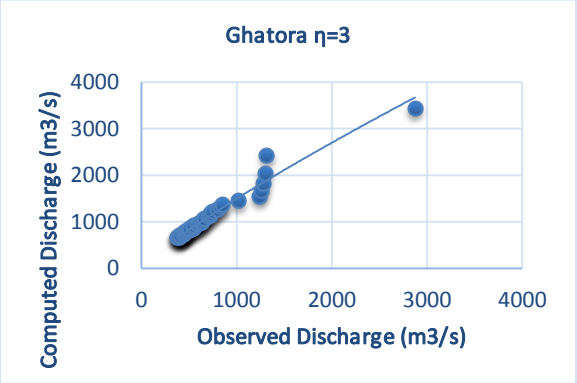

Basantpur η=2.5
Computed Discharge (m3/s)
40000
30000
20000
10000
0
0
10000
20000
30000
40000
Observed Discharge (m3/s)

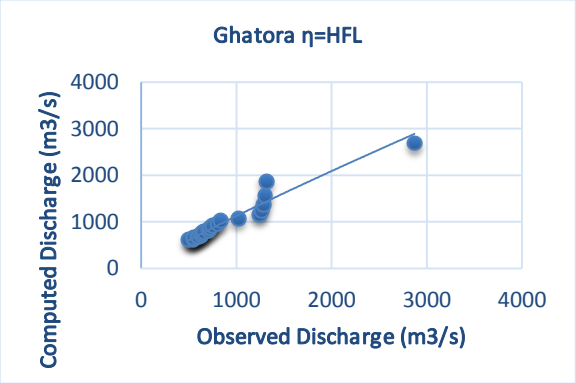

Basantpur η=3
Computed Discharge (m3/s)
40000
30000
20000
10000
0
0
10000
20000
30000
40000
Observed Discharge (m3/s)

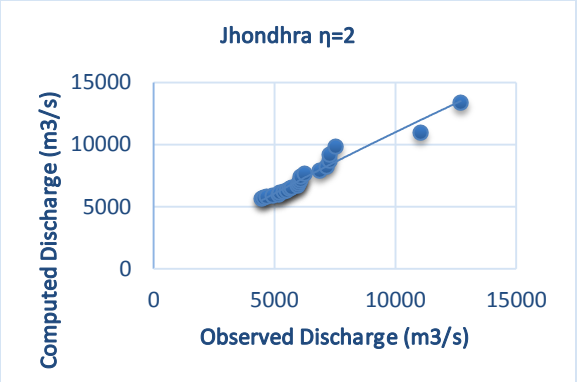

Basantpur η=HFL
Computed Discharge (m3/s)
40000
30000
20000
10000
0
0
10000
20000
30000
40000
Observed Discharge (m3/s)

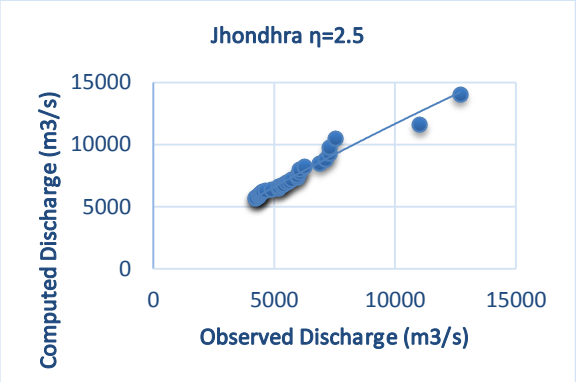

Ghatora η=2
Computed Discharge (m3/s)
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Observed Discharge (m3/s)

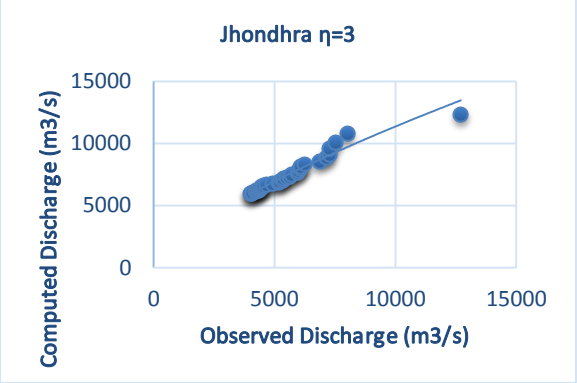

Ghatora η=2.5
Computed Discharge (m3/s)
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Observed Discharge (m3/s)

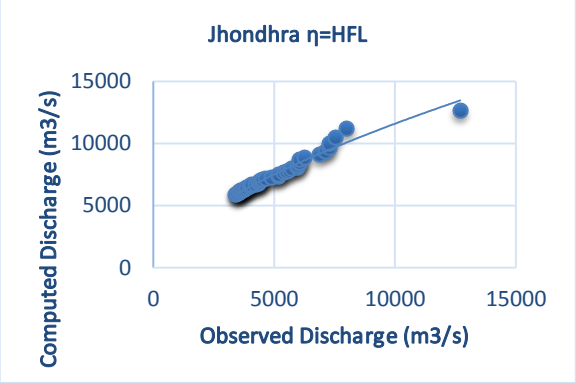

Ghatora η=3
Computed Discharge (m3/s)
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Observed Discharge (m3/s)

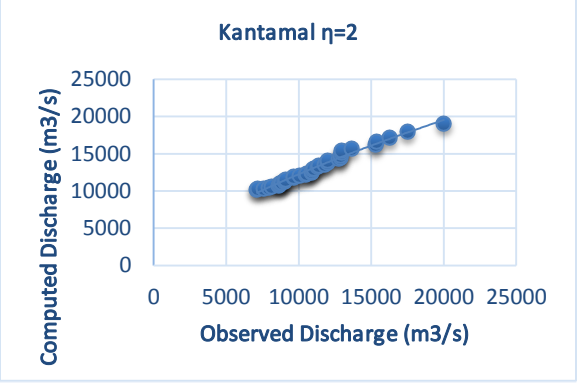

Ghatora η=HFL
Computed Discharge (m3/s)
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Observed Discharge (m3/s)

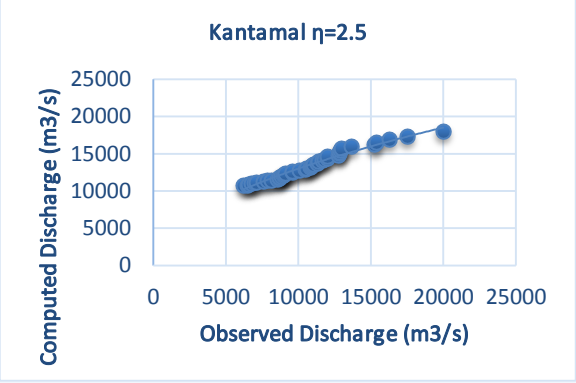

Jhondhra η=2
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000

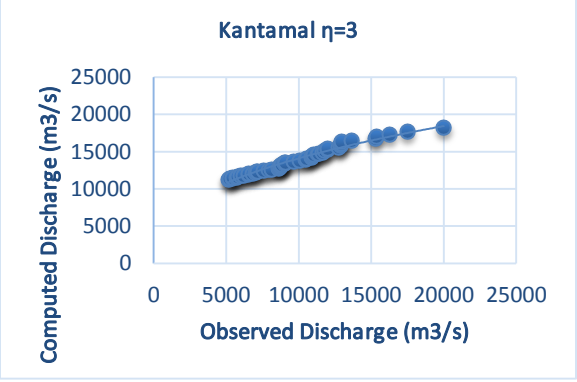

Jhondhra η=2.5
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000

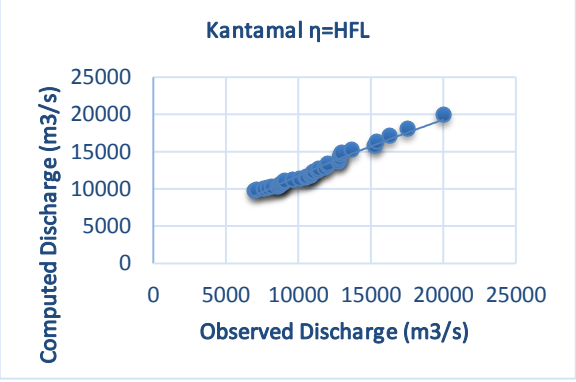

Jhondhra η=3
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000

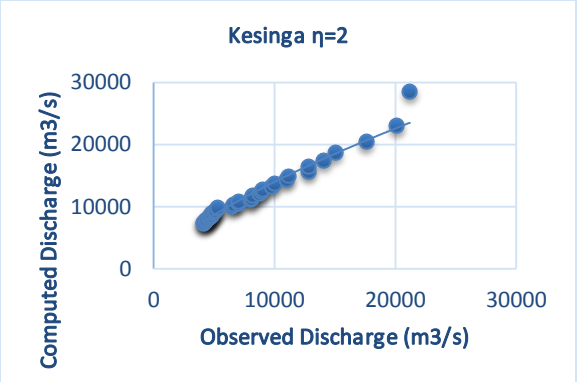

Jhondhra η=HFL
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000

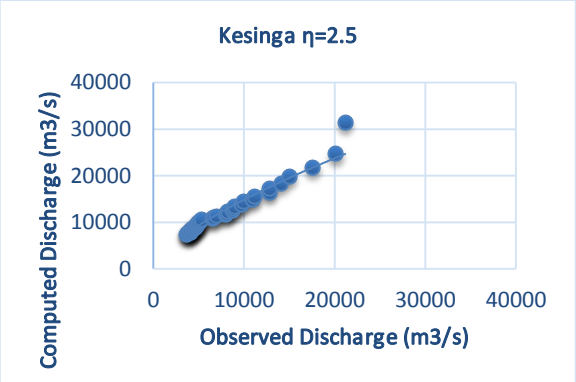

Kantamal η=2
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000
25000

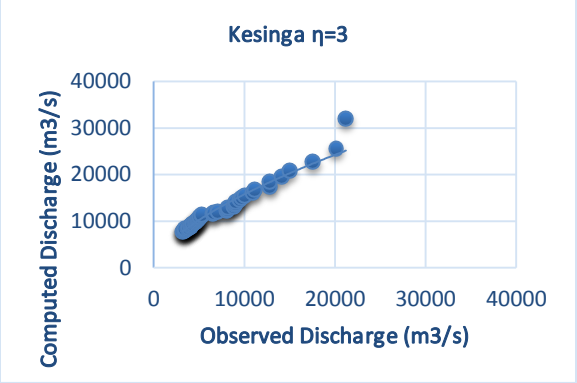

Kantamal η=2.5
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000
25000

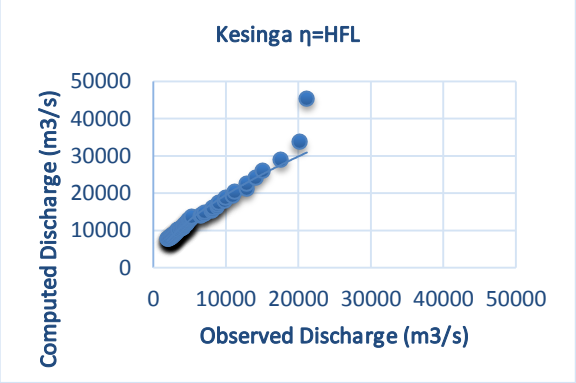

Kantamal η=3
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000
25000

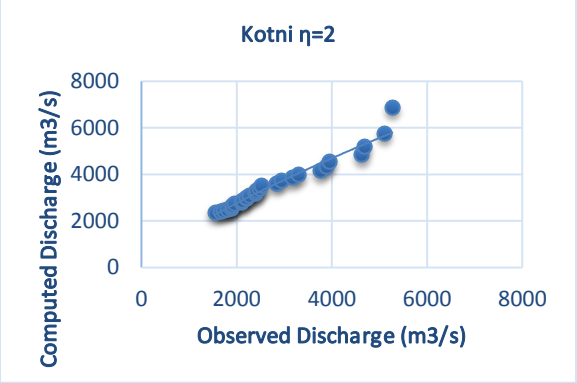

Kantamal η=HFL
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000
25000

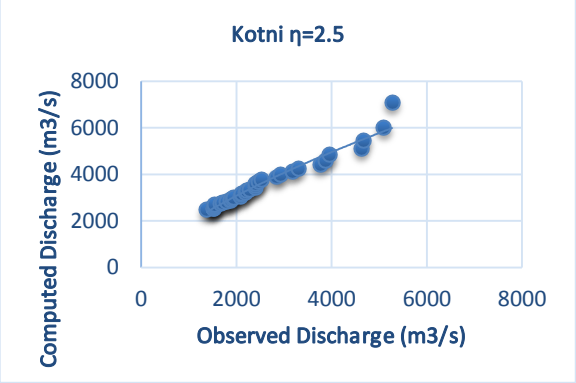

Kesinga η=2
Computed Discharge (m3/s)
30000
20000
10000
0
0
10000
20000
30000
Observed Discharge (m3/s)

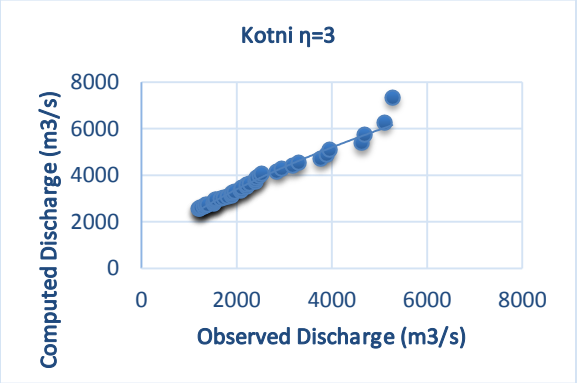

Kesinga η=2.5
Computed Discharge (m3/s)
40000
30000
20000
10000
0
0
10000
20000
30000
40000
Observed Discharge (m3/s)

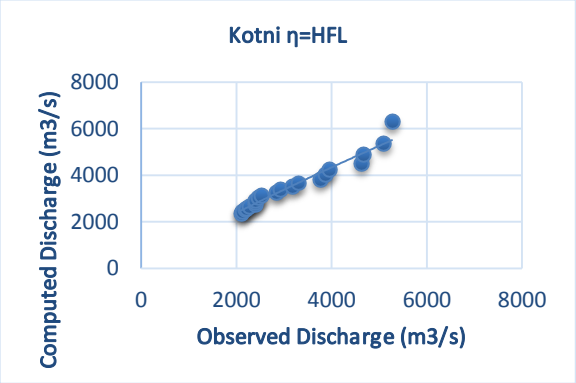

Kesinga η=3
Computed Discharge (m3/s)
40000
30000
20000
10000
0
0
10000
20000
30000
40000
Observed Discharge (m3/s)

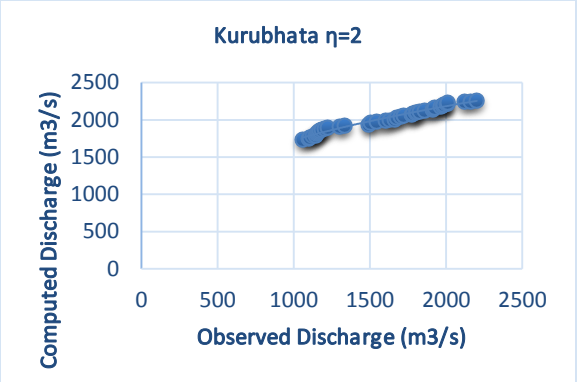

Kesinga η=HFL
Computed Discharge (m3/s)
50000
40000
30000
20000
10000
0
0
10000
20000
30000
40000
50000
Observed Discharge (m3/s)

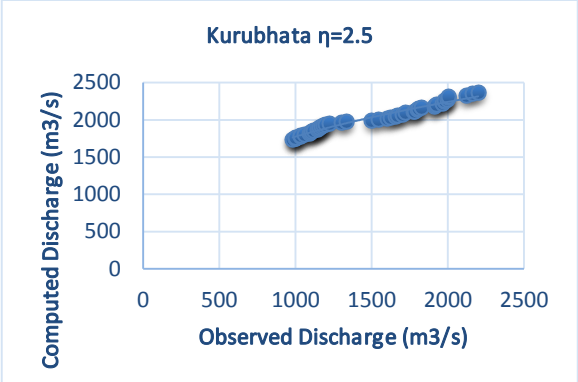

Kotni η=2
Computed Discharge (m3/s)
8000
6000
4000
2000
0
0
2000
4000
6000
8000
Observed Discharge (m3/s)

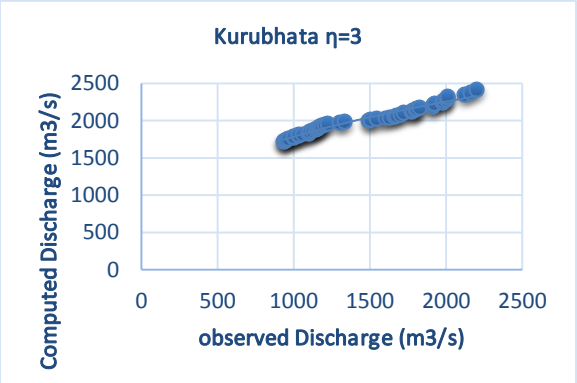

Kotni η=2.5
Computed Discharge (m3/s)
8000
6000
4000
2000
0
0
2000
4000
6000
8000
Observed Discharge (m3/s)

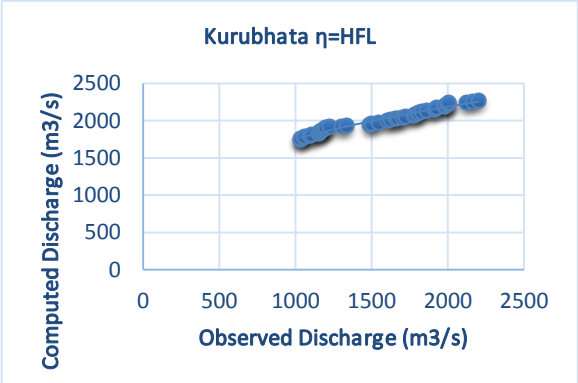

Kotni η=3
Computed Discharge (m3/s)
8000
6000
4000
2000
0
0
2000
4000
6000
8000
Observed Discharge (m3/s)

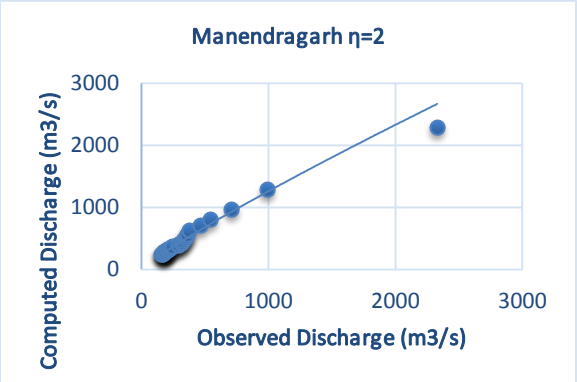

Kotni η=HFL
Computed Discharge (m3/s)
8000
6000
4000
2000
0
0
2000
4000
6000
8000
Observed Discharge (m3/s)

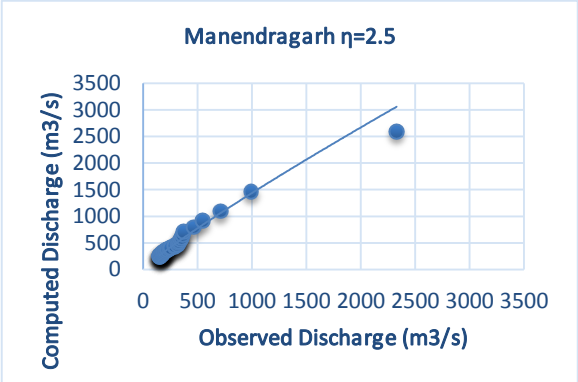

Kurubhata η=2
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
500
1000
1500
2000
2500

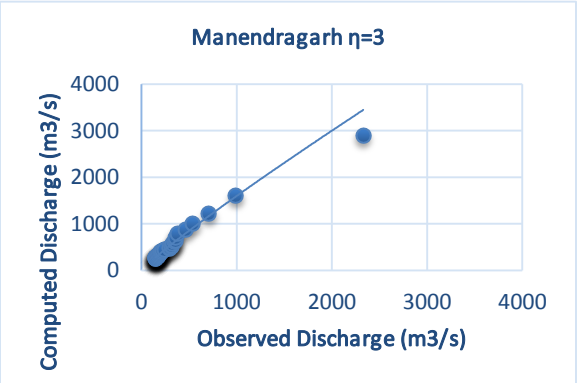

Kurubhata η=2.5
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
500
1000
1500
2000
2500

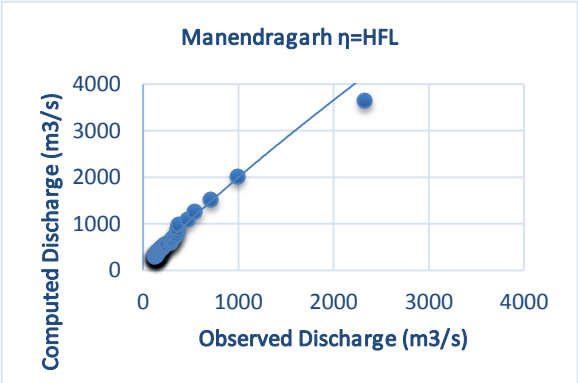

Kurubhata η=3
Computed Discharge (m3/s)
observed Discharge (m3/s)
0
500
1000
1500
2000
2500

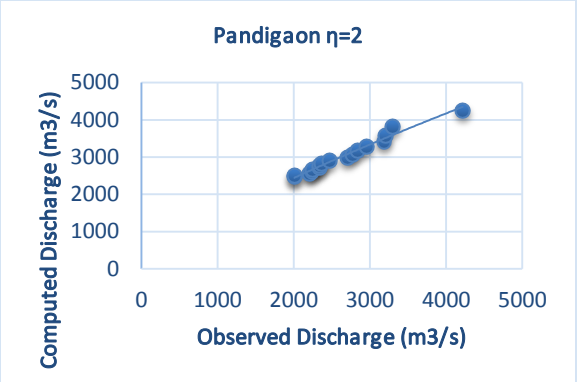

Kurubhata η=HFL
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
500
1000
1500
2000
2500

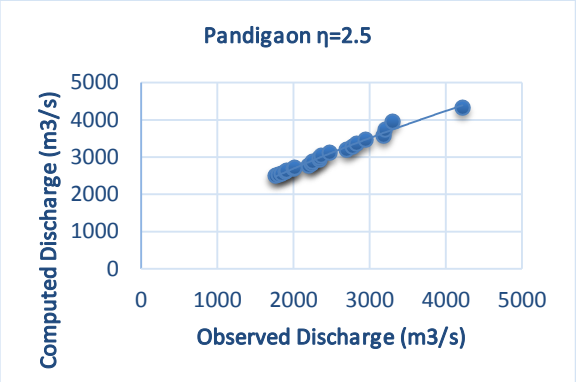

Manendragarh η=2
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
1000
2000
3000

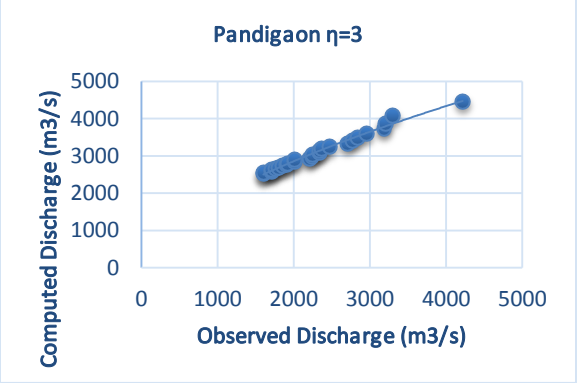

Manendragarh η=2.5
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
500
1000
1500
2000
2500
3000
3500

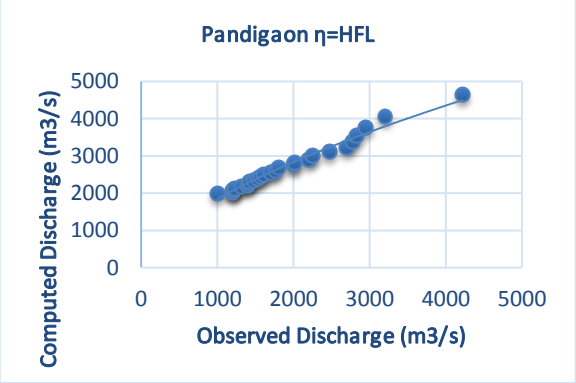

Manendragarh η=3
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
1000
2000
3000
4000

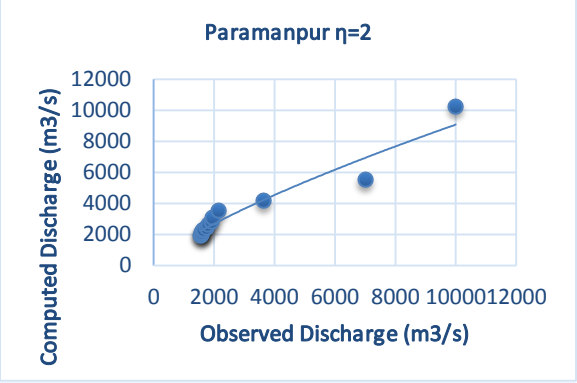

Manendragarh η=HFL
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
1000
2000
3000
4000

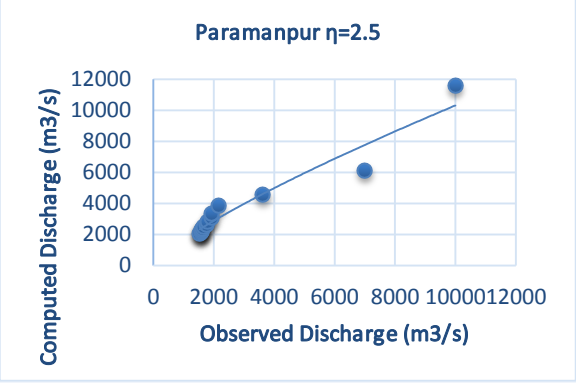

Pandigaon η=2
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
1000
2000
3000
4000
5000

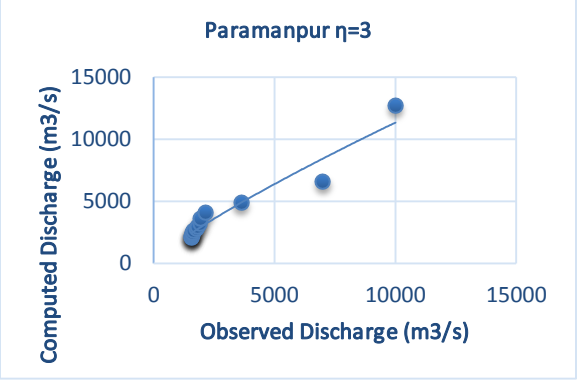

Pandigaon η=2.5
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
1000
2000
3000
4000
5000

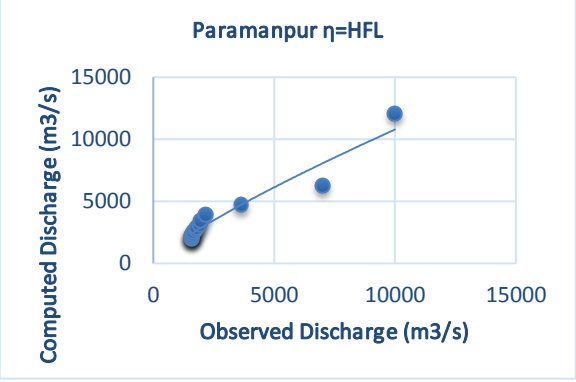

Pandigaon η=3
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
1000
2000
3000
4000
5000

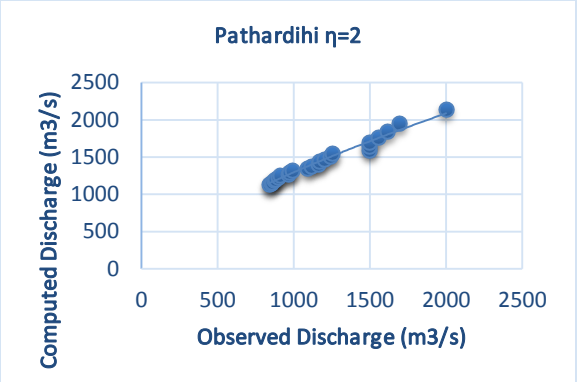

Pandigaon η=HFL
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
1000
2000
3000
4000
5000

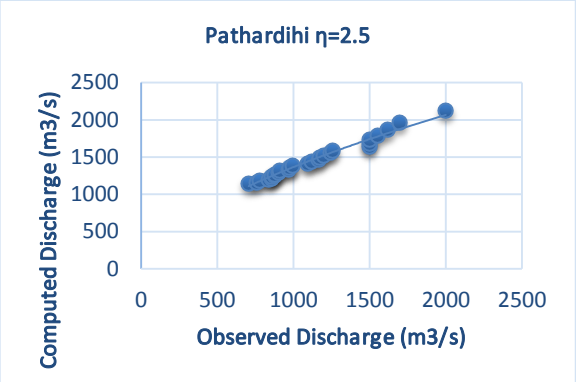

Paramanpur η=2
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
2000
4000
6000
8000
10000
12000

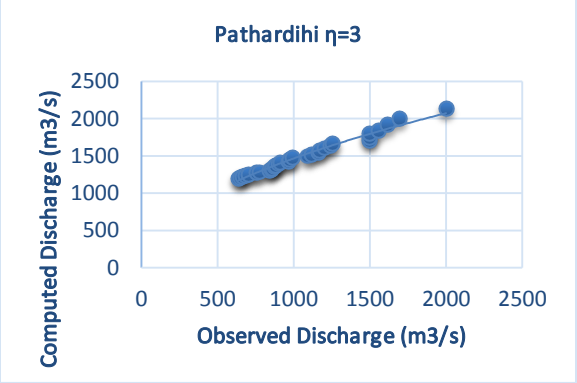

Paramanpur η=2.5
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
2000
4000
6000
8000
10000
12000

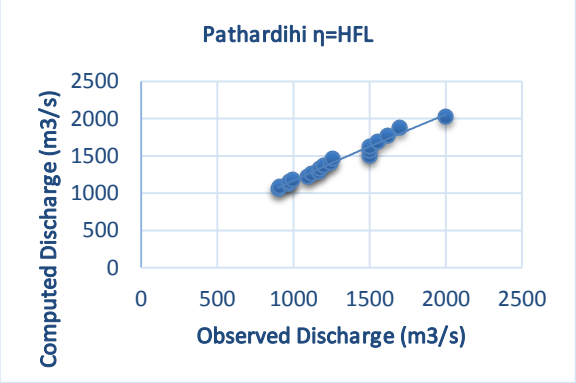

Paramanpur η=3
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000

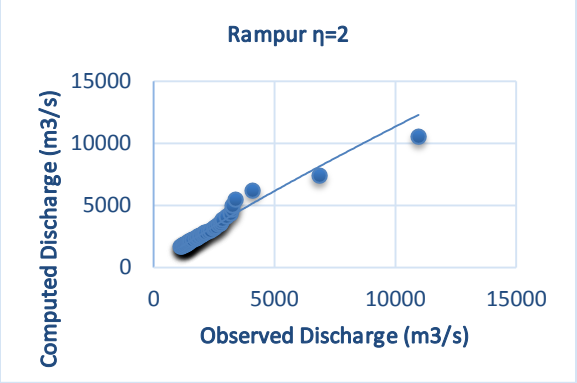

Paramanpur η=HFL
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000

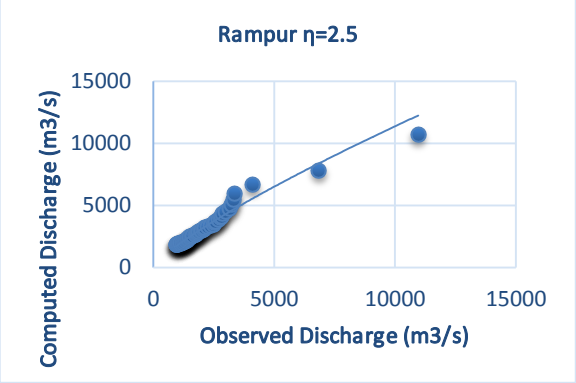

Pathardihi η=2
Computed Discharge (m3/s)
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2500
Observed Discharge (m3/s)

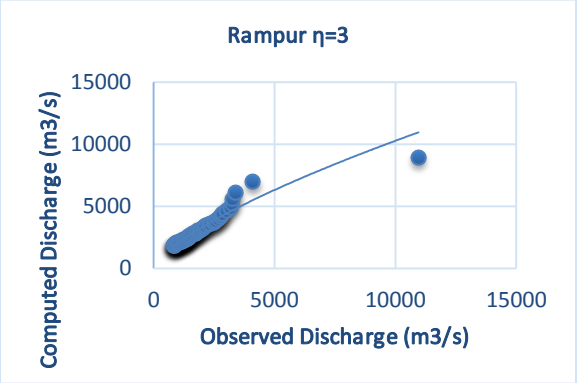

Pathardihi η=2.5
Computed Discharge (m3/s)
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2500
Observed Discharge (m3/s)

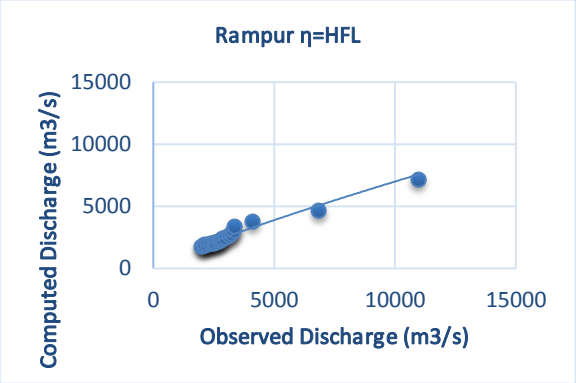

Pathardihi η=3
Computed Discharge (m3/s)
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2500
Observed Discharge (m3/s)

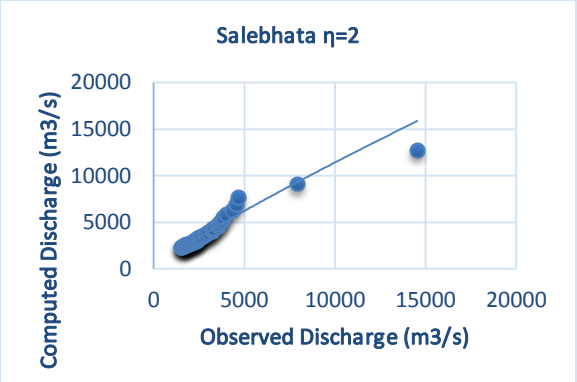

Pathardihi η=HFL
Computed Discharge (m3/s)
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2500
Observed Discharge (m3/s)

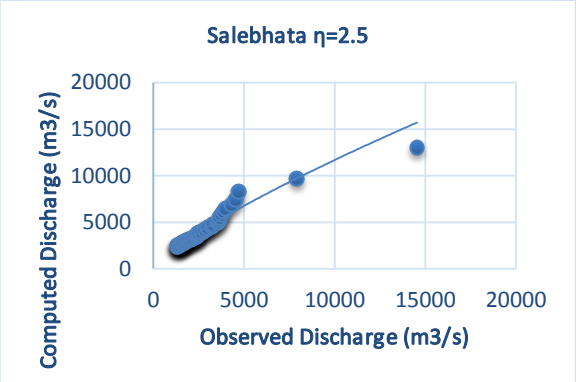

Rampur η=2
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

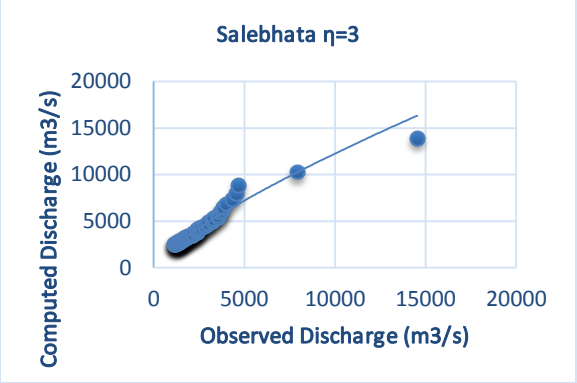

Rampur η=2.5
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

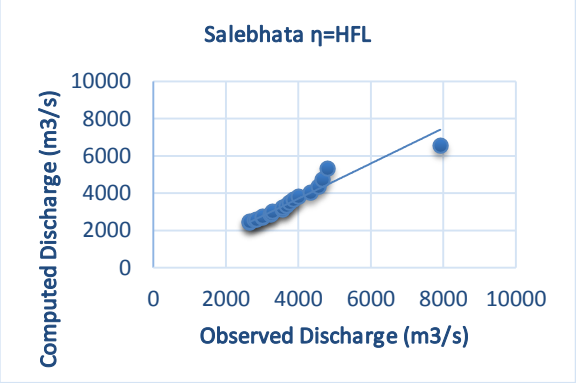

Rampur η=3
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

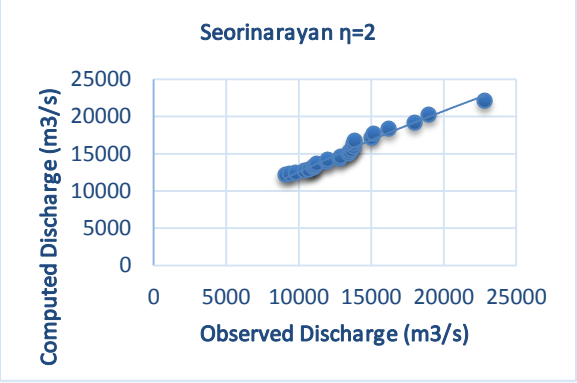

Rampur η=HFL
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

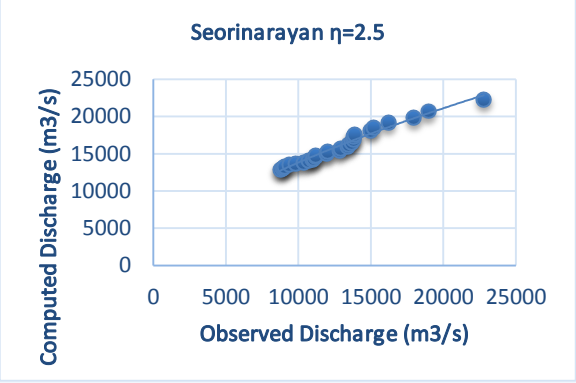

Salebhata η=2
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000

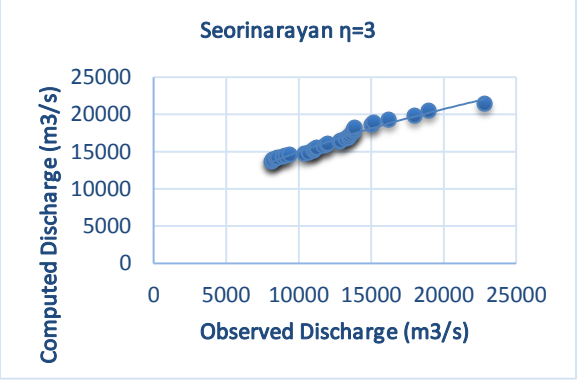

Salebhata η=2.5
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000

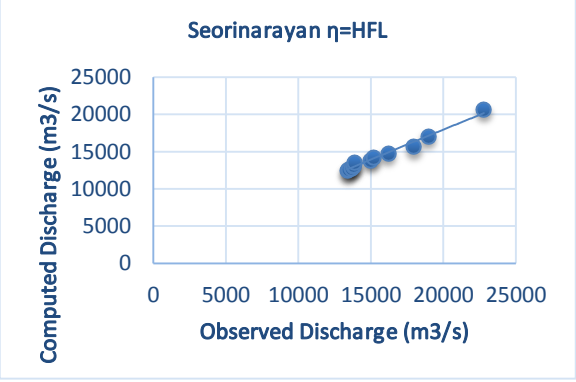

Salebhata η=3
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000

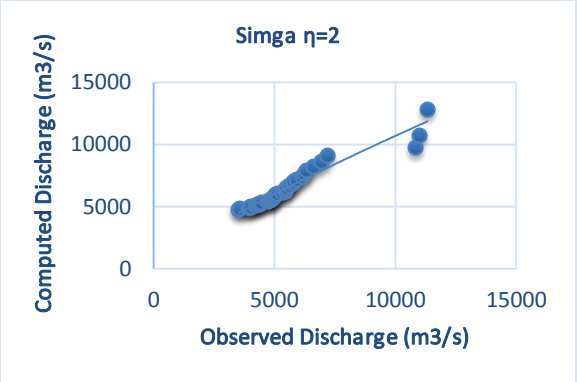

Salebhata η=HFL
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
2000
4000
6000
8000
10000

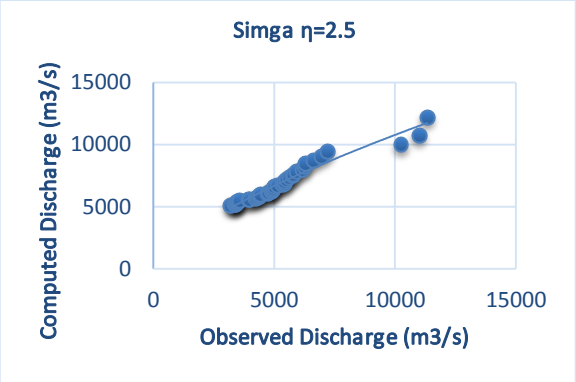

Seorinarayan η=2
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000
25000

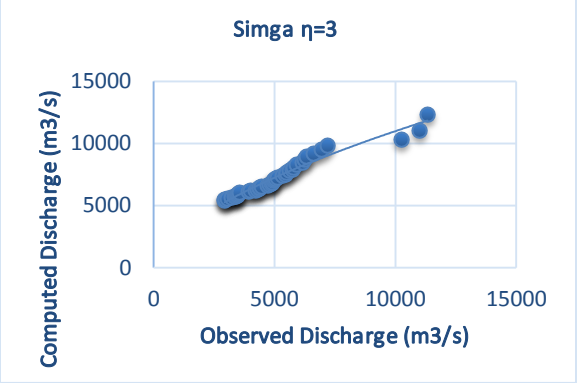

Seorinarayan η=2.5
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000
25000

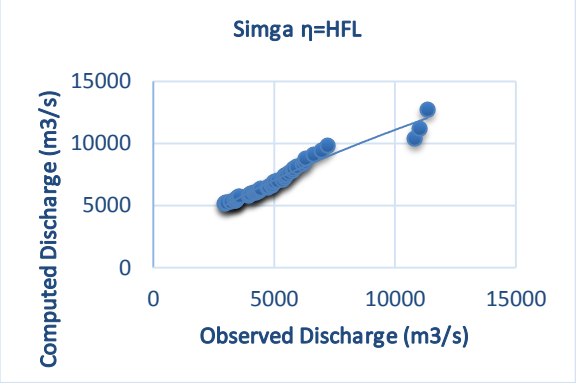

Seorinarayan η=3
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000
25000

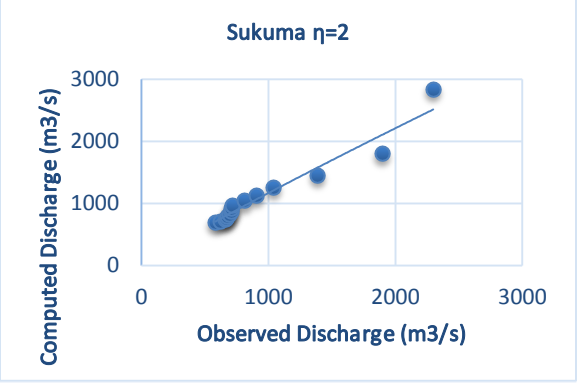

Seorinarayan η=HFL
Computed Discharge (m3/s)
Observed Discharge (m3/s)
0
5000
10000
15000
20000
25000

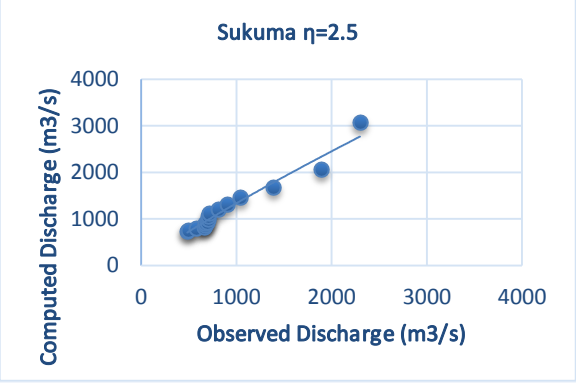

Simga η=2
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

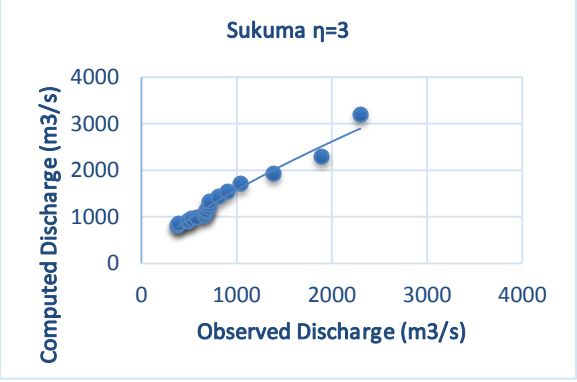

Simga η=2.5
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

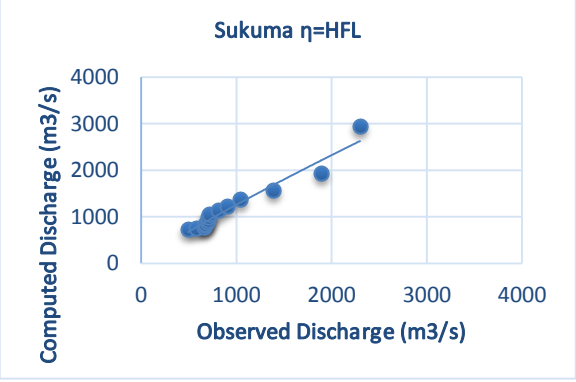

Simga η=3
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)

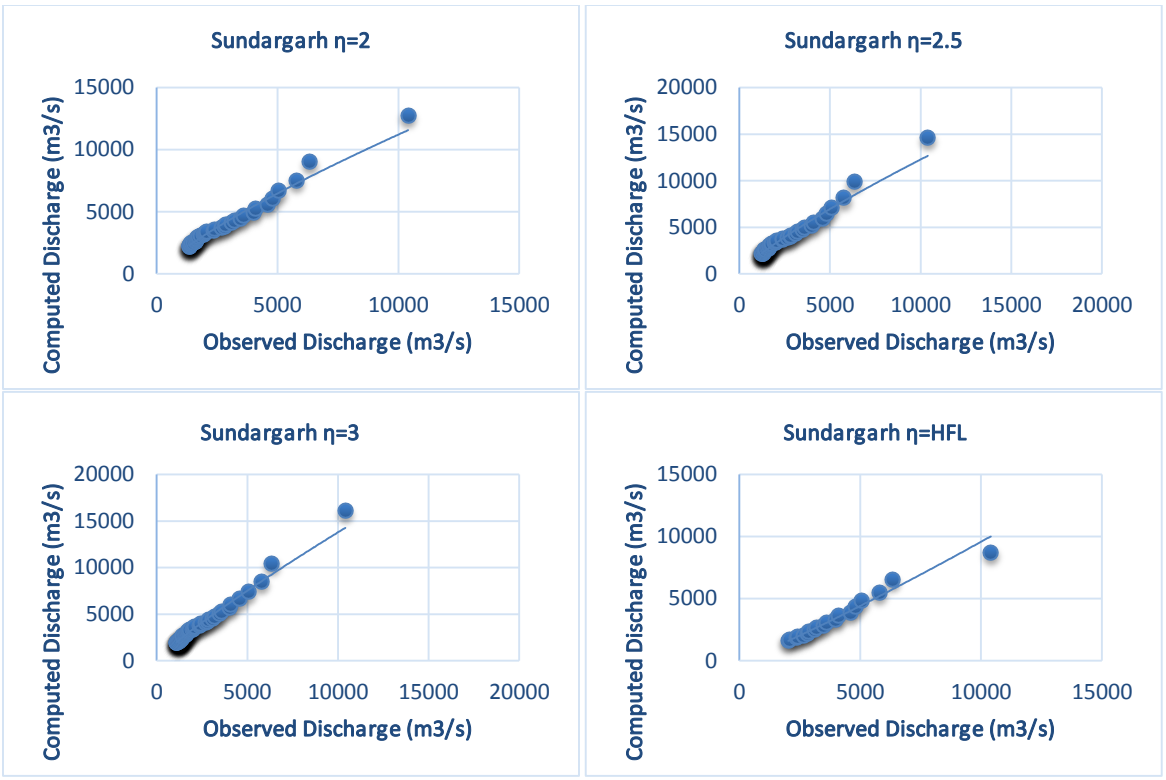

Simga η=HFL
Computed Discharge (m3/s)
15000
10000
5000
0
0
5000
10000
15000
Observed Discharge (m3/s)


Sukuma η=2
Computed Discharge (m3/s)
3000
2000
1000
0
0
1000
2000
3000
Observed Discharge (m3/s)


Sukuma η=2.5
Computed Discharge (m3/s)
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Observed Discharge (m3/s)


Sukuma η=3
Computed Discharge (m3/s)
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Observed Discharge (m3/s)


Sukuma η=HFL
Computed Discharge (m3/s)
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Observed Discharge (m3/s)



QQ Plots at Tikarapara and Rajim Station for a range of thresholds

APPENDIX II

List of Variables of different Stations

STATION	CA	AVG SLOPE	DD	CR	PBF	PBB	AMP	ELEVATION	LSL	PBA	PERIMETER	PSCL	PSL	FF	SF
ANDHIYARKORE	0.075	0.611	0.194	0.508	0.596	0.471	0.847	0.886	0.225	0.469	0.202	0.205	0.310	0.176	0.456
BAMNIDHI	0.298	0.764	0.172	0.516	0.487	0.071	0.914	0.792	0.508	0.594	0.410	0.088	0.422	0.137	0.587
BARONDA	0.108	0.604	0.157	0.553	0.989	0.002	0.885	0.865	0.289	0.125	0.265	0.099	0.415	0.154	0.519
BASANTPUR	0.042	0.172	0.160	0.476	0.086	0.149	0.914	0.418	0.211	0.918	0.142	0.036	0.452	0.113	0.712
GHATORA	0.099	0.852	0.153	0.484	0.596	0.009	0.536	0.822	0.271	0.499	0.222	0.062	0.441	0.160	0.502
JONDHRA	0.262	0.321	0.171	0.450	0.368	0.232	0.536	0.731	0.427	0.586	0.336	0.041	0.479	0.171	0.470
KANTAMAL	0.286	0.999	0.160	0.559	0.770	0.013	0.866	0.808	0.372	0.369	0.435	0.415	0.110	0.245	0.327
KESINGA	0.211	0.853	0.152	0.609	0.619	0.022	0.966	0.678	0.326	0.505	0.407	0.445	0.596	0.236	0.341
KOTNI	0.238	0.245	0.161	0.434	0.303	0.003	0.847	0.564	0.332	0.780	0.308	0.544	0.097	0.256	0.314
KURUBHATA	0.159	0.926	0.161	0.527	0.568	0.011	0.958	0.821	0.383	0.572	0.306	0.222	0.318	0.129	0.623
SUNDERGARH	0.198	0.712	0.150	0.484	0.383	0.040	0.991	0.918	0.365	0.745	0.314	0.178	0.344	0.177	0.454
SUKUMA	0.042	0.589	0.150	0.453	0.252	0	0.91	0.635	0.167	0.864	0.136	0.504	0.095	0.182	0.441
SIMGA	0.252	0.270	0.187	0.497	0.279	0.933	0.885	0.670	0.357	0.666	0.363	0.079	0.429	0.234	0.343
SEORINARAYAN	0.215	0.251	0.159	0.670	0.274	0.071	0.914	0.507	0.459	0.800	0.452	0.064	0.432	0.121	0.663
SALEBHAT	0.157	0.525	0.154	0.486	0.202	0.065	0.911	0.531	0.368	0.904	0.280	0.255	0.191	0.137	0.584

RAMPUR	0.116	0.426	0.159	0.560	0.398	0.029	0.885	0.596	0.391	0.719	0.277	0.372	0.212	0.090	0.891
PATHARDIHI	0.083	0.104	0.158	0.447	0.066	0.024	0.885	0.473	0.282	0.989	0.188	0	0.507	0.124	0.645
PARAMNPUR	0.012	0.284	0.175	0.409	0.334	0	0.984	0.441	0.082	0.805	0.067	0	0.426	0.222	0.361
PANDIGAON	0.192	0.929	0.168	0.518	0.745	0.003	0.966	0.718	0.339	0.390	0.331	0.328	0.213	0.198	0.404
MANENDRAGARH	0.033	0.584	0.168	0.468	0.526	0	0.799	0.998	0.142	0.563	0.124	0.507	0.122	0.195	0.411
RAJIM	0.179	0.478	0.160	0.489	0.614	0.005	0.885	0.749	0.369	0.458	0.301	0	0.488	0.156	0.515
TIKRAPADA	0.999	0.607	0.163	0.686	0.428	0.065	0.899	0.558	0.997	0.657	0.999	0.162	0.254	0.119	0.674

PUBLISHED PAPERS AND PAPERS UNDER REVIEW

- Guru N, and Jha R., "Flood estimation in Mahanadi river system, India using partial duration series" *Journal of Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*.
- Guru N, and Jha R., "Flood frequency analysis for Tel sub-basin of Mahanadi River, India using Weibull, Gringorten and L-moments formula" *International Journal of Inovative Research and Creative Technology*.
- Guru N, and Jha R., "Comparison between Annual Maximum and Peaks Over Threshold Models for Flood Frequency Analysis for Tel Subbasin of Mahanadi, Odisha" *Journal of Water Resource and Protection. (Accepted)*
- Guru N, and Jha R., "A Framework for the Selection of Threshold in Partial Duration Series Modeling" *Journal of Hydrology. (Under Review)*
- Guru N, and Jha R., "Partial duration series and its implication on regional frequency analysis-case study of Mahanadi basin" *Journal of Hydrologic Engineering. (Under Review)*
- Guru N, and Jha R., "Application of Soft Computing Techniques for River Flow Prediction in the Downstream Catchment of Mahanadi River Basin Using Partial Duration Series, India" *Journal of Stochastic Environmental Research and Risk Assessment. (Under Review)*
- Guru N, and Jha R., "At site Flood Frequency Analysis for the Tel Subbasin" *Proceedings of the Symposium on Sustainable Infrastructure Development (SID) 8th-9th February 2013, IIT Bhubaneswar, Bhubaneswar, Odisha, India.*
- Guru N, and Jha R., "Comparison between Annual maximum and Peaks over threshold models for Flood Frequency Analysis for Tel Subbasin of Mahanadi, Odisha" *Proceedings of the International Workshop on Terrestrial Water Cycle Observation and Modeling from Space: Innovation and Reliability of Data Products (WATGLOBS Beijing, China) 26-30 April, 2013.*
- Guru N, and Jha R., "Maximum Likelihood Parameter estimators for the Tel subbasin of Mahanadi, Odisha" *Proceedings of the National Conference on Recent Advancement in Civil Engineering, NIT Patna, Bihar, India, June 13-14 2013.*

- Guru N, and Jha R., "Criteria for selection of model for Flood frequency analysis" *Proceedings of the Interational Conference on Hydraulics, Water Resources, Coastal and Environmental Engineering (HYDRO-2013), IIT Madras,Chennai, India.*
- Guru N, and Jha R., "A Study on selection of probability distributions for at-site flood frequency analysis in Mahanadi river basin, India" *Proceedings of the International conference on Fluvial Hydraulics, Lausanne, Switzerland 3-5 Sept, 2014.*
- Guru N, and Jha R., "Flood Frequency analysis for Tel Subbasin of Mahanadi River India using Weibull, Gringorten and L-moments formula" *International conference on Modelling tools for Sustainable Water Resources Management in IIT, Hyderabad, Telangana, India, Dec 2014.*
- Guru N, and Jha R., "Flood Frequency Analysis of Tel Basin of Mahanadi River System, India using Annual Maximum and POT Flood Data" *Proceedings of the International conference on Water Resources, Coastal and Ocean Engineering, NIT, Suratkal, Karnataka, India, 12-15 March 2015.*
- Guru N, and Jha R., "Estimating Threshold Values for Partial Duration Series of Peak Flows in Mahanadi River Basin, India" *International Water Conference 2016 (IWC2016) on "Water Resources in Arid Area: the Way Forward*". (Accepted)