FLOOD ESTIMATION AND FORECASTING IN MAHANADI RIVER BASIN USING SOFT COMPUTING TECHNIQUES

A THESIS

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

of

DOCTOR OF PHILOSOPHY

in

HYDROLOGY

by

ANIL KUMAR KAR

DEPARTMENT OF HYDROLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)

AUGUST, 2011

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **FLOOD ESTIMATION AND FORECASTING IN MAHANADI RIVER BASIN USING SOFT COMPUTING TECHNIQUES** in partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy** and submitted in the Department of Hydrology, Indian Institute of Technology, Roorkee is an authentic record of my work carried out during a period from August, 2008 to August, 2011 under the supervision of Dr. N. K. Goel, Professor, Department of Hydrology, Indian Institute of Technology Roorkee and Dr. Anil Kumar Lohani, Scientist, National Institute of Hydrology, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree or diploma of this or any other Institute.

(ANIL KUMAR KAR)

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

(N. K. Goel)
Supervisor

(A. K. Lohani)
Supervisor

Date: 25/8/2011

The Ph.D. Viva-voce examination of Mr. Anil Kumar Kar, Research Scholar, has been held on Oct 31, 2011

C. Chatterjee

Signature of Supervisor(s)

Signature of External Examiner

Chairman SRC

Professor & Head

ABSTRACT

GENERAL ABOUT THE PROBLEM

Water is known as the most precious gift of nature for growth of civilization as well as a destructive element causing mass devastation. Flood hazards have become ever increasing natural disasters resulting in the highest economic damage among all kinds of natural disasters around the world. The country India is full of rivers and rainfall patterns are heavily influenced by monsoon. Thus occurrence of flood remains an inevitable feature in most parts of the country. The large river systems like Ganga, Brahmaputra, Godavari and Mahanadi influence the flood scenario of the country. Mahanadi is the 6th largest river system in India. The river is also known for its huge water potential and frequent flood devastations.

Chhatisgarh and Orissa states of India cover almost 99% of the catchment area of Mahanadi basin. Currently a number of developmental projects are going on in these two states. For these projects well defined flood estimate formulae are required. The lower reach of Mahanadi basin is in the state of Orissa and flood is a permanent threat to this reach. Hirakud reservoir is the only major flood controlling structure in the basin. The downstream area of Hirakud is around 58000 km^2. It remains uncontrolled and experiences frequent floods. Flood damages can be reduced drastically by adopting various non-structural measures such as flood frequency prediction and flood forecasting.

In the present study efforts have been made to develop regional flood formulae for the entire Mahanadi basin using L-moment and prioritized variables based approach. For the lower reach of Mahanadi basin (downstream of Hirakud dam) flood forecasting models have been developed using soft computing techniques like ANN and Fuzzy logic. The performance of soft computing models has been compared with conventional and conceptual models.

BROAD OBJECTIVES

In the present study the flood problem of Mahanadi basin has been addressed by developing regional flood formulae for the uncontrolled portion of the basin and by developing a flood forecasting model using ANN and fuzzy logic for the lower reach. The objectives are summarized as follows:

i. Development of regional flood formulae for Mahanadi basin.

ii. Development of a flood forecasting model for the reach downstream of Hirakud, and
iii. Development of a key raingauge network for Kantamal sub-basin of lower Mahanadi basin for flood forecasting.

DEVELOPMENT OF REGIONAL FLOOD FORMULAE FOR MAHANADI BASIN

Mahanadi basin has been divided into homogeneous regions by applying different clustering techniques, as the entire basin is not hydro-meteorologically homogeneous. Principal component analysis (PCA) has been used to prioritise the site characteristics. Clustering techniques like Hierarchical Clustering (HC), K-mean (KM), Fuzzy C-mean (FCM), Kohonen Self Organization Map (SOM) and Andrews Plot (AP) are applied on prioritised variables to verify the results of clustering. The entire basin is divided into two homogeneous clusters based on the results of Fuzzy C-mean (FCM) technique. L-moment based methods are used to test the homogeneity of the clusters and to identify a best fitting underlying frequency distribution. The Generalised Pareto (GP) distribution holds good for cluster-1 and it contains the areas which can contribute substantially towards runoff generation due to high slope and drainage density characteristics. The cluster-2 contains areas with low runoff generation capacity as compared to cluster-1. Generalised Extreme Value (GEV) is the robust distribution for this cluster.

DEVELOPMENT OF A FLOOD FORECASTING MODEL FOR THE REACH DOWNSTREAM OF HIRAKUD

In this reach the discharge data are available at 3 hour interval during monsoon period. However due to various reasons, sometimes only peak discharge data are available. Three hourly and peak discharge data of three different gauge and discharge (G&D) stations located downstream of Hirakud dam have been used to develop various flood forecasting models based on soft computing methods like Multi Layered Feed Forward-Artificial Neural Network (MLFF-ANN), Radial Basis Function-Artificial Neural Network (RBF-ANN) and Takagi Sugeno (TS)-Fuzzy inference system. The forecasting results of these models are compared with statistical and time lag methods exercised by the Department of Water Resources, Government of Orissa. The TS-fuzzy models perform better than other models for peak flood as well as 3-hour flood forecasting. The TS-fuzzy model gives an efficiency of 86.7% for a lead time of 42-hours. This is expected to improve the existing operational flood forecasting system quite significantly.

DEVELOPMENT OF A KEY RAINGAUGE NETWORK FOR KANTAMAL SUB-BASIN OF LOWER MAHANADI BASIN FOR FLOOD FORECASTING

Sometimes it becomes difficult to collect data from all the rain gauges either due to instrumental disorder, difficulty in transmission, inability to take readings and many other operational difficulties during flood times. Therefore, the network of key raingauge stations is designed. The performance of key rain gauge network in flood forecasting is discussed and demonstrated through a case study of Kantamal sub-catchment of Mahanadi basin. This sub-catchment significantly contributes to the downstream floods at Khairmal, Barmul and Mundali. The Fuzzy logic applied on the network developed through AHP has shown the best result for flood forecasting at Kantamal gauge and discharge site with efficiency of 82.74% and RMSE value of 500.2 m^3/s for 1-day lead period forecast. Analytic Hierarchy Process (AHP) has been successfully introduced for the first time in this study for establishing the key rain gauge network.

The contents of these have been presented in seven chapters namely (i) Introduction, (ii) Description of study area and data used, (iii) Review of literature, (iv) Regional flood frequency analysis of Mahanadi basin using prioritized variables, (v) Development of flood forecasting model for downstream of Hirakud reservoir, (vi) Rain gauge network design of Kantamal sub-basin of lower Mahanadi for flood forecasting, and (vii) Conclusions and scope for further work.

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to my supervisors, Dr. N.K. Goel, Professor, Department of Hydrology, Indian Institute of Technology, Roorkee and Dr. A. K. Lohani, Scientist, National Institute of Hydrology, Roorkee. Their wide knowledge and logical way of thinking have been of great value for me. Their understanding, encouraging and personal guidance have provided a good basis for the present thesis. I am thankful to Dr. D.K. Srivastava, Chairman, Student Research Committee (SRC), Dr. M. Perumal, member, SRC, Dr. S.K. Mishra, member, SRC and Dr. Himanshu Joshi, Professor and Head, Department of Hydrology for useful suggestions during the course of this study.

I express my sincere gratitude to Prof. B. S. Mathur, Prof. Ranvir Singh, Prof. D. C. Singhal, Dr. D.S. Arya and Dr. M.K. Jain, faculty members Department of Hydrology, for their constant encouragement and valuable suggestions during the course of this research work. I am also thankful to Dr. Senthil Kumar, Dr. Sanjay Jain, Dr. Manohar Arora, Dr. P.C. Nayak, Dr. R.P. Pandey and Mr. Ajeet Singh Chhabra of National Institute of Hydrology, Roorkee for useful interactions.

I warmly thank Mr. C.S. Padhi, Mr. R. M. Das, Mr. H. Mohanty, Mr. M. S. Rao, Mr. Deepak Jhajharia, Mr. M.K. Chaudhury, Mr. Maskey, Mr. Vinit Jain, Mr. Gosavi Eknath, Miss Asmita and Mr. Negash, Research Scholars, Department of Hydrology for their valuable advice and timely help.

I am also thankful to Mr. Iliyas, Mr. Jamshed, Mr. Amarjeet Singh, Mr. Raj Kumar, Mr. Neeraj Sharma, Mr. Sagar, Mr. Bhanesh, Mr. Navneet, Mr. Rohit, Mr. Shyam and all other support staff of Department for their help.

I also admire the sincere efforts of Er. Harischandra Behera, Er. G.P. Roy, Er. Bighnaraj Purohit, Dr. B.K. Sethi, Dr. P.K. Parhi and other officers of Government of Orissa for extending all technical and administrative help from time to time.

I am thankful to my father esteemed **Sri Upendra Nath Kar** and mother **Smt. Kanak Prava Kar** for their continuous support. They have been the constant source of inspiration in my life.

A lot of time which should have been devoted towards my wife **Sunita** and daughter **Suman** and **Adwiti** was devoted to this work. I am thankful to them for allowing me to do so patiently and with full affection.

The greatest of my gratitude is towards the almighty **Lord Jagannath** to whom I bow with reverence for giving me the power to fulfill the promises he entrusted to me as a human being.

Last but not the least my sincere thanks are due to all those who helped me directly or indirectly to complete this research work successfully.

Anil Kumar

Place: Roorkee **(Anil Kumar Kar)**

Date: 25.08.11

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF NOTATIONS / ABBREVIATIONS

Symbols having a common meaning are defined here. Other locally used symbols are defined wherever the same have been used.	
ANN	Artificial Neural Network
Avg.	Average
BM	Barmul-Mundali
C_k	Coefficient of Kurtosis
C_s	Coefficient of Skewness
C_v	Coefficient of Variation
CWC	Central Water Commission
DEM	Digital Elevation Module
DOWR	Department of Water Resources
Eff.	Nash-Sutcliffe Efficiency
GEV	Generalised Extreme Value distribution
GL	Generalised Logistics distribution
GN	Generalised Normal distribution
GP	Generalised Pareto distribution
KB	Khairmal-Barmul
KBM	Khairmal-Barmul-Mundali
KM	Khairmal-Mundali
Max.	Maximum
Min.	Minimum
MLFF	Multi- Layered Feed Forward
MLP	Multi-Layered Perceptron
OLS	Orthogonal Least Square
PT-III	Pearson Type-III distribution
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SRTM	Shuttle Radar Topography Mission

St. dev.	Standard Deviation
T	Return period
TS	Takagi Sugeno
Q_m	Mean Annual Peak Flood
Q_T	T-years return period flood

1.1 GENERAL ABOUT THE PROBLEM

Water is known as the most precious gift of nature for growth of civilization as well as a destructive element causing mass devastation. Flood hazards have become ever increasing natural disasters resulting in the highest economic damage among all kinds of natural disasters around the world. The country India is full of rivers and rainfall patterns are heavily influenced by monsoon. Thus occurrence of flood remains an inevitable feature in most parts of the country. The large river systems like Ganga, Brahmaputra, Godavari and Mahanadi influence the flood scenario of the country. Mahanadi is the 6^{th} largest river system in India. The river is also known for its huge water potential and frequent flood devastations.

Chhatisgarh and Orissa states of India cover almost 99% of the catchment area of Mahanadi basin. Currently a number of developmental projects are going on in these two states. For these projects well defined flood estimate formulae are required. The lower reach of Mahanadi basin is in the state of Orissa and flood is a permanent threat to this reach. Hirakud reservoir is the only major flood controlling structure in the basin. The area downstream of Hirakud is around 58000 km^2. It remains uncontrolled and experiences frequent floods. Flood damages can be reduced drastically by adopting various non-structural measures such as flood frequency prediction and flood forecasting.

In the present study efforts have been made to develop regional flood formulae for the entire Mahanadi basin using L-moment and prioritized variables based approach. For the lower reach of Mahanadi basin (downstream of Hirakud dam) flood forecasting models have been developed using soft computing techniques like ANN and Fuzzy logic. The performance of soft computing models has been compared with conventional and conceptual models.

1.2 BROAD OBJECTIVES

In the present study the flood problem of Mahanadi basin has been addressed by developing regional flood formulae for the uncontrolled portion of the basin and by developing a flood forecasting model using ANN and fuzzy logic for the lower reach. The objectives are summarized as follows:

i. Development of regional flood formulae for Mahanadi basin.
ii. Development of a flood forecasting model for the reach downstream of Hirakud, and

iii. Development of a key raingauge network for Kantamal sub-basin of lower Mahanadi basin for flood forecasting.

1.3 DEVELOPMENT OF REGIONAL FLOOD FORMULAE FOR MAHANADI BASIN

Mahanadi basin has been divided into homogeneous regions by applying different clustering techniques, as the entire basin is not hydro-meteorologically homogeneous. Principal component analysis (PCA) has been used to prioritise the site characteristics. Clustering techniques like Hierarchical Clustering (HC), K-mean (KM), Fuzzy C-mean (FCM), Kohonen Self Organization Map (SOM) and Andrews Plot (AP) are applied on prioritised variables to verify the results of clustering. The entire basin is divided into two homogeneous clusters based on the results of Fuzzy C-mean (FCM) technique. L-moment based methods are used to test the homogeneity of the clusters and to identify a best fitting underlying frequency distribution.

1.4 DEVELOPMENT OF A FLOOD FORECASTING MODEL FOR DOWNSTREAM OF HIRAKUD

Three hourly and peak discharge data of three different gauge and discharge (G&D) stations located downstream of Hirakud dam have been used to develop various flood forecasting models based on soft computing methods like Multi Layered Feed Forward-Artificial Neural Network (MLFF-ANN), Radial Basis Function-Artificial Neural Network (RBF-ANN) and Takagi Sugeno (TS)-Fuzzy inference system. The forecasting results of these models are compared with statistical and time lag methods exercised by the Department of Water Resources, Government of Orissa.

1.5 DEVELOPMENT OF A KEY RAINGAUGE NETWORK FOR KANTAMAL SUB-BASIN OF LOWER MAHANADI BASIN FOR FLOOD FORECASTING

Sometimes it becomes difficult to collect data from all the rain gauges either due to instrumental disorder, difficulty in transmission, inability to take readings and many other operational difficulties during flood times. Therefore, the network of key raingauge stations is designed. The performance of key rain gauge network in flood forecasting is discussed and demonstrated through a case study of Kantamal sub-catchment of Mahanadi basin. This sub-catchment significantly contributes to the downstream floods at Khairmal, Barmul and Mundali.

1.6 LAYOUT OF THE THESIS

The work has been organized in the form of seven chapters as follows:

Chapter 1 introduces the research work, its need and briefly describes the objectives of the study. The details of study area and data used in the study are presented in **Chapter 2**. **Chapter 3** presents a review of the flooding problems related to Mahanadi basin, historical development of the flood control measures taken in the basin, hydrological studies related to this basin, existing flood estimation formulae and flood forecasting method. A brief review of hydrological application of soft computing techniques is also presented in this chapter. Clustering techniques using prioritised variables to delineate homogeneous regions are discussed in **Chapter 4**. The results of regional flood frequency analysis using L moments are also presented in this chapter. The details of flood forecasting models using ANN and Fuzzy logic are given in **Chapter 5**. The results of comparison with statistical and time lag methods are also given in this chapter. Multi-criteria decision analysis and clustering methods to establish key raingauge network have been discussed in **Chapter 6**. The application of the selected network in flood forecasting has also been demonstrated in this chapter. **Chapter 7** summarises the conclusions drawn in the present study and the scope for further work in the area.

2.1 BACKGROUND

This chapter gives details of study area and data used in carrying out the study. As the study comprises of flood prediction, flood forecasting and design of rain gauge network, the collection of data varies in a wide range but all the data are confined to Mahanadi basin only. The basin comprising a catchment area of 141569 km^2 is the 6th largest basin of India and situated at its eastern part. The basin lies within the geographical co-ordinates of 80^0-30' to 86^0-50' of East Longitude and 19^0-20' to 23^0-35' of North Latitude (Fig. 2.1). India is divided into 7 major zones, which are further subdivided into 26 hydro-meteorologically homogeneous sub zones. Mahanadi basin occupies a major part of sub-zone 3(d). The chapter presents the description of Mahanadi river basin, and details of hydro-meteorological data used in the study.

2.2 MAHANADI RIVER BASIN

The Mahanadi is a major east flowing river in peninsular river system. It originates near Pharasiya village of Raipur district of Chhatisgarh state and covers 851 km. of length before falling to Bay of Bengal at Orissa coast. It is an interstate river covering Chhatisgarh, Orissa, Jharkhand, Madhya Pradesh and Maharastra. The basin is physically bounded by Central Hills in the North, Eastern Ghat in the South, Maikala hill range in the West and by Bay of Bengal in the East. The basin is largely divided into four parts viz. Central table land, Northern plateau, Eastern ghats and Coastal plain.

The basin is comprised of very vast area and there is a large variation in geographic and climatic conditions throughout the basin. The state wise coverage of drainage area of the river Mahanadi is shown in table 2.1.

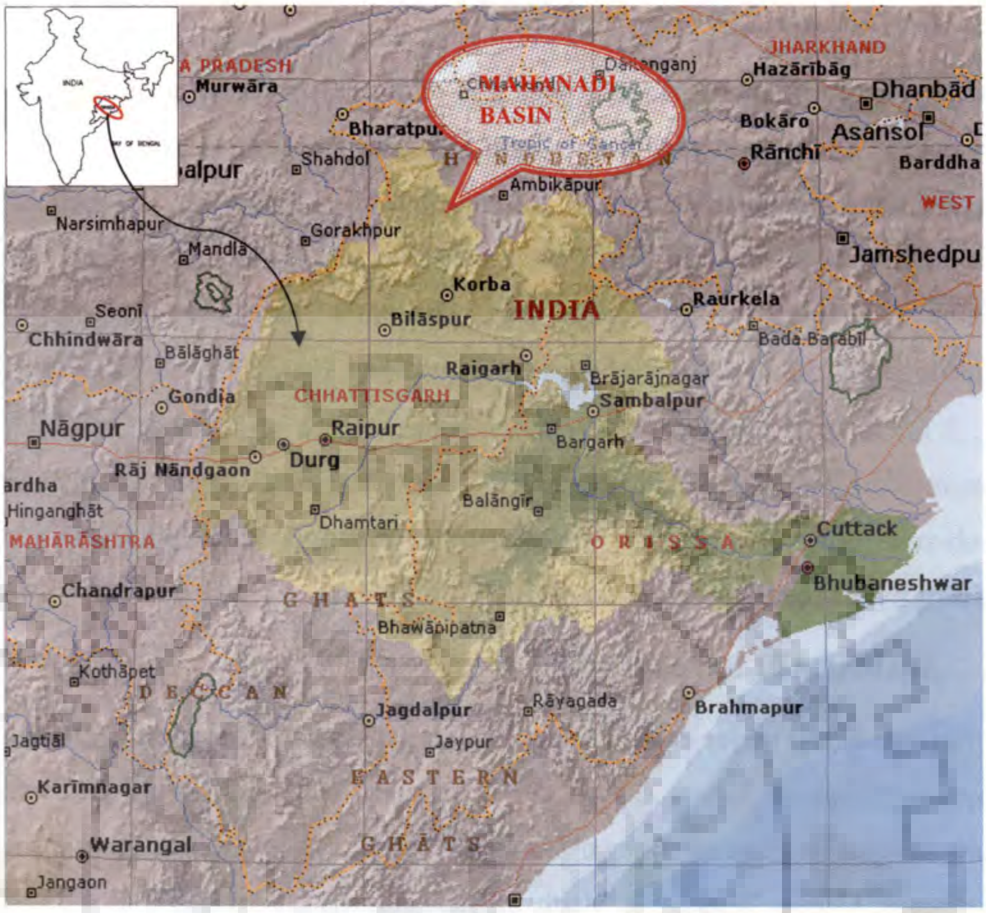


Fig. 2.1 Location map of Mahanadi basin

(Source: http://Encarta.msn.com/map_701605802/Mahanadi _Basin.html.)

Table 2.1 State wise distribution of Mahanadi basin

Sl. No.	State	Catchments Area (km^2)	Percentage to total basin
1	Madhya Pradesh and Chhatisgarh	75,336	53.21
2	Orissa	65889	46.53
3	Maharastra	238	0.17
4	Jharkhand	126	0.09
	Total	1,41,589	100

Fourteen major tributaries of river Mahanadi are Seonath, Hasdeo, Mand, Kelo, Birai, Pairi, Jonk, Sukha, Kanki, Lialr, Lath, Ong, Tel and Ib (NWDA, 2004). The water resource potential of the basin is around 48732 million cubic meters (75% dependability). The river Mahanadi shares major part of subzone-3(d). Total area covered under this zone is 1,95,256 km^2. Besides Mahanadi two other basins Brahmani and Baitarani are the parts of this zone.

Figure 2.2 shows location of this subzone over all hydro-meteorological subzones. Table 2.2 shows the catchment areas of various rivers in subzone-3(d).

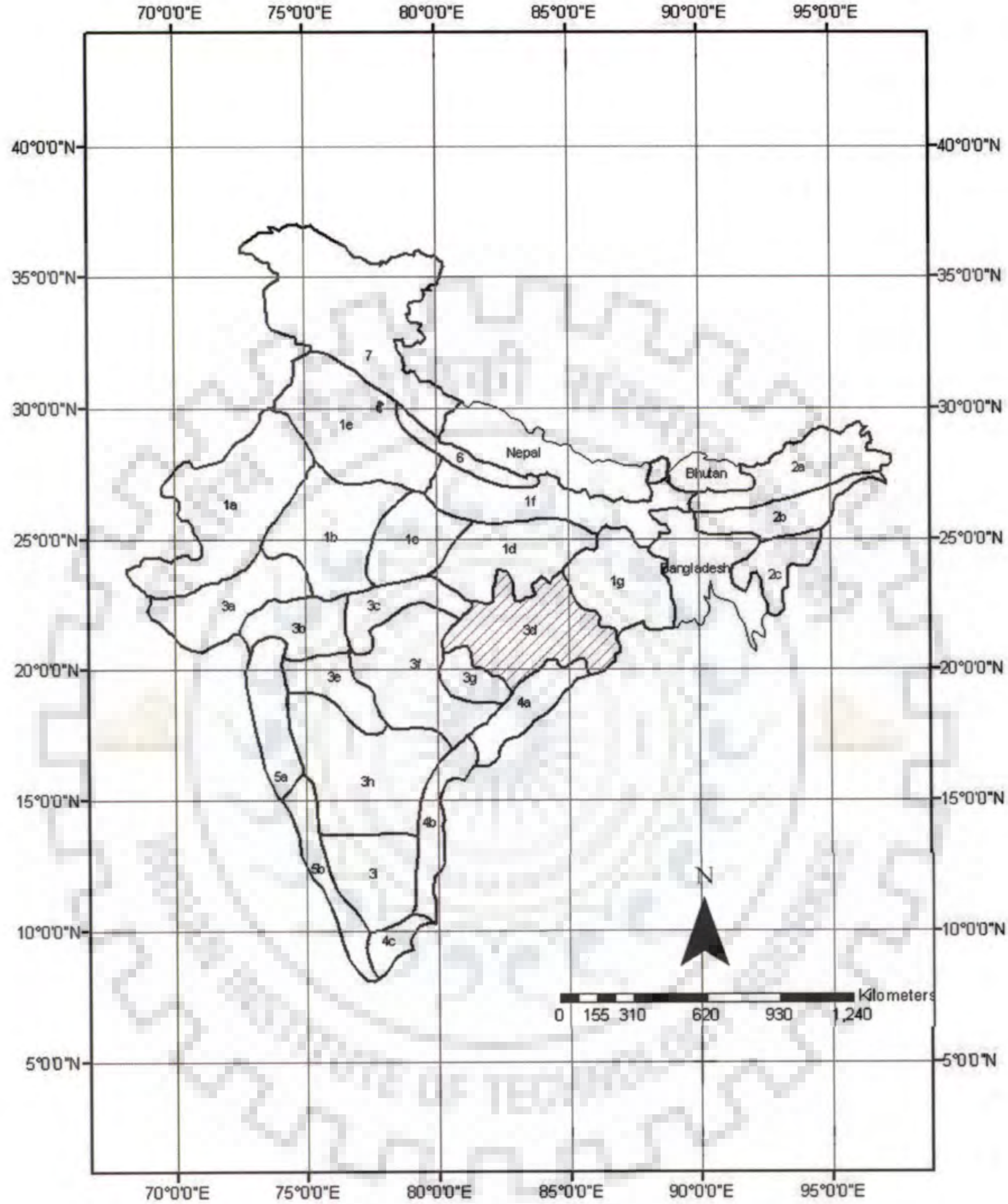


Fig. 2.2 Hydro-meteorological sub-zones of India

Table 2.2 Basinwise details of subzone-3(d)

Basin name	Catchment area (km^2)	Stream length (km)
Mahanadi	1,40,628	850
Brahmani	35,337	705
Baitarani	19,291	333
Total	**1,95,256**	

Source: NIH (1997)

The salient characteristics of Mahanadi basin are described in the subsequent sections.

2.2.1 Basin shape

The basin is roughly circular in shape with a diameter of about 400 km. and exit passage is 60 km. wide and 160 km. long. The basin has a Horton form factor of 0.66.

2.2.2 Topography

The upper reach of the basin lies in a very undulating plateau with hillocks eroded moulds. The southern part of the plateau is open but to the east and west there are a number of hill ranges which have steep slopes resulting in water draining directly into the Mahanadi river. The basin continuously slopes towards the main valley with no congestion.

2.2.3 Climate

The climate in the basin area is tropical monsoon type with distinct seasons viz. Summer from March to May, the monsoon season from June to September and winter from October to February. The hottest and coldest months of the year are May and December respectively.

2.2.4 Rainfall

The basin receives about 90% of its rainfall during the monsoon season. Generally, the southwest monsoon sets by the middle of the June over the entire basin and remains active till the end of September. The spatial variation in rainfall is moderate in the basin. The formation of depressions in the Bay of Bengal cause cyclones which bring about wide spread heavy rains resulting in floods and destructions. The basin falls in the south-west monsoon track thus receives heavy rainfall during monsoon periods.

2.2.5 Temperature

The coldest and hottest months in the sub-basin are December and May respectively. The highest monthly mean maximum temperature is 42.1^0C while lowest monthly mean temperature is 8.2^{0}C. The highest single point temperature is 47.7^{0}C and lowest is 0^{0}C. The maximum temperature occurs at central table land whereas lowest temperature in Eastern ghat is 8.7 ^{0}C.

For development of regional flood formulae entire Mahanadi basin has been considered. For the flood prone lower reach of Mahanadi a flood forecasting system has been developed. Key rain gauge network has been developed for Tel sub-catchment of the lower Mahanadi. Further details of these study areas are given in individual chapters.

2.3 DATA USED IN THE STUDY

This study ranges from development of regional flood formulae, to development of flood forecasting model and development of a key rain gauge network. Hence for these studies different parts of Mahanadi have been used, as explained in chapter 1. The frequency and length and other details of various data used in the study are shown in Table 2.3. Most of these data were collected from DOWR, Govt. of Orissa and CWC, New Delhi. The schematic diagram of Mahanadi basin is shown in Fig. 2.3.

Table 2.3 Inventory of data used in this study

Sl. No.	Station	Data type	Source	Frequency	Period
1	20 G&D sites of Mahanadi basin	Discharge	DOWR, Orissa/ CWC, New Delhi	Daily	1971-2007 (Varying between 9 to 37 years among different G&D sites)
2	Khairmal, Barmul, Mundali,	Discharge	DOWR, Orissa	03 hourly	1996 – 2010
3	14 Raingauge sites of Kantamal sub-basin	Rainfall	DOWR, Orissa	Daily	2000-2005
4	Dasapalla	Evaporation	DOWR, Orissa	Daily	2000 - 2005
Derived data					
Sl. No.	**Area**	**Data type**			**Source**
1	Mahanadi basin	Catchment area, slope, drainage density, longest stream length			90 m SRTM data
2	Mahanadi basin	Normal rainfall, Maximum 1-day rainfall for gauge and discharge sites			DOWR, Orissa
3	Kantamal sub-basin	Daily average rainfall, Maximum 1-day rainfall			DOWR, Orissa

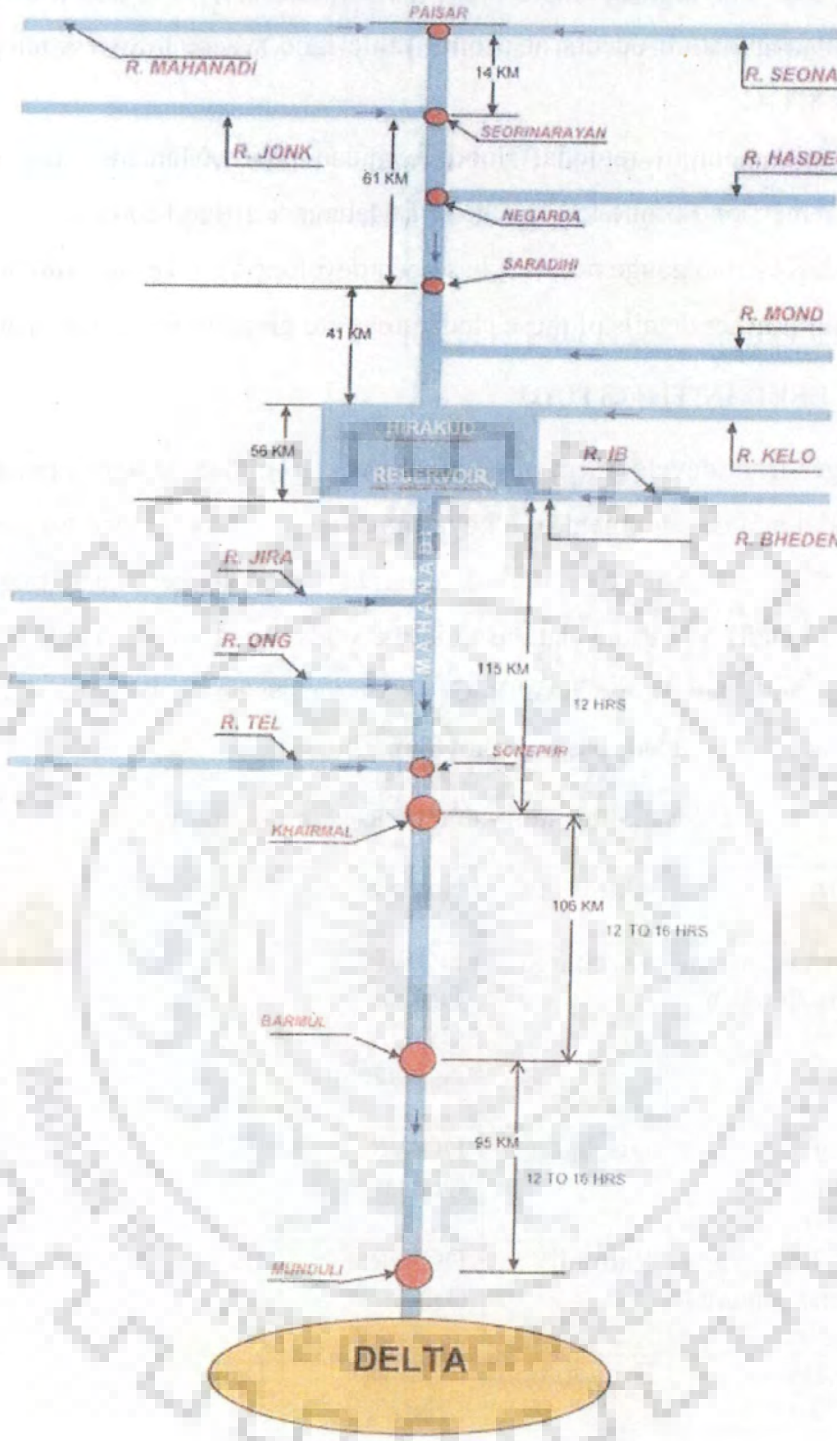


Fig. 2.3 Schematic diagram of Mahanadi basin

2.4 DATA PROCESSING

A number of operations as listed below were applied on the raw data, before these were used for analysis.

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

The Excel and MATLAB software were used for the processing of data.

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

2.5 DATA STATISTICS

The statistical parameters of annul maximum series (AMS) used for regional flood frequency analysis are shown in Table 2.4. The peak flood data are derived from the available 3-hourly discharge data. The statistical parameters of floods data at Khairmal, Barmul and Mundali are shown in Table 2.5 and 2.6 per peaks and 3-hourly discharge data respectively. The daily rainfall data (2000 – 2005) of 14 rain gauge stations of Kantamal sub-basin were collected, and 29 isolated storms were identified from these data. The statistical parameters of daily rainfall data of 14 rain gauge stations are tabulated in Table 2.7 for calibration period (2000 – 2003) and validation period (2004 – 2005). For Kantamal sub-basin, the discharge data are observed at Kantamal G&D site and evaporation data are observed at Dasapalla. The missing values are filled using arithmetic mean method and outliers are checked using Grubb-Beck method. The statistical parameters of discharge data and evaporation data are tabulated in Table 2.8 and 2.9 respectively.

Table 2.4 Statistical parameters of Annual Maximum Series (in m^3/s) of 20 G&D sites of Mahanadi basin

Station Name	Max.	Min.	Avg.	St. dev.	C_v	C_s	C_k
Sundergarh	10404	800.3	2393.8	2101.3	0.88	2.48	10.01
Kurubhata	2160	756	1459.9	450.1	0.31	-0.04	2.06
Ghatora	2281	137.2	788.5	530.9	0.67	1.38	5.58
Jondhra	12700	1600	5190.6	2497	0.48	1.28	5.64
Basantapur	33087.9	2441	13189.2	6534.2	0.5	1.03	4.75
Andhiyarkor	990	46.7	332.6	216.7	0.65	1.56	5.65
Bamanidhi	9583.1	758.9	3183	2129.1	0.67	1.02	3.9
Rampur	10958.4	90	1973.4	1867.6	0.95	3.26	17.93
Salebhata	14545	114	2505.4	2565.4	1.02	3.31	16.8
Baronda	6593.8	205.5	2199.1	2060.8	0.94	0.95	2.82
Rajim	8620.2	255.6	3793.8	2964.2	0.78	0.39	1.74
Kotni	5269	463.3	1777.4	1166.6	0.66	1.5	6.54
Simga	11332	858.3	4601.2	2625.1	0.57	0.99	4.24
Kantamal	17500	1180	8150.9	4782.4	0.59	0.16	2.08
Kesinga	5269	463.3	1777.4	1166.6	0.66	1.5	6.54
Kelo	1612.6	125.8	741.1	448.2	0.6	0.84	3.55
Mahendragarh	21192	600	6877.2	5237.7	0.76	1	3.91
Pandigaon	2088	97.5	464.8	491.6	1.06	2.73	11.37
Pathardih	4217	1211	2618.6	981.6	0.37	-0.02	4.19
Sukuma	2315	35	841.1	730	0.87	0.96	3.42

Table 2.5 Statistical parameters of peak flow (m^3/s) at three stations

	Calibration(1996 – 2003)			Validation (2004 – 2010)		
	Khairmal	Barmul	Mundali	Khairmal	Barmul	Mundali
Max.	36300.41	42223.60	39868.98	41035.00	41960	44742.30
Min.	2660.20	2992.73	2841.32	2760.20	3035.18	3062.06
Avg.	13267.61	14627.51	15061.74	11682.81	12126.52	14457.75
St.dev.	8694.41	10031.97	10190.82	7417.93	7593.20	8604.68
C_v	0.66	0.69	0.68	0.63	0.63	0.60
C_s	0.97	0.90	0.86	1.69	1.01	1.13
C_k	0.53	0.19	0.04	4.64	1.01	2.14

Table 2.6 Statistical parameters of 3-hourly flow (m^3/s) at three stations

	Calibration (1996 – 2003)			Validation (2004 – 2010)		
	Khairmal	Barmul	Mundali	Khairmal	Barmul	Mundali
Max.	36300.41	42223.60	39869.00	41035.00	41960.00	44742.30
Min.	266.00	346.70	291.50	464.10	624.00	707.50
Avg.	3748.98	4580.34	4476.47	4126.91	4474.96	5328.16
St.dev.	4805.48	5510.48	5800.18	4565.61	4864.91	5868.90
C_v	1.28	1.20	1.30	1.11	1.09	1.10
C_s	2.80	2.79	3.00	2.65	2.62	2.31
C_k	9.75	9.07	10.35	10.38	8.76	6.94

Table 2.7 Statistical parameters of daily rainfall data (mm) of Kantamal sub-basin

Station Id	RG station name	Max.	Min.	Avg.	St. dev.	C_v	C_s	C_k
Calibration (2000 – 2003)								
KI8	Bhaskel	247.4	0	9.59	23.53	2.45	4.59	28.83
KI9	Kurumuli	525	0	11.15	32.56	2.92	8.98	119.68
M25	Sagada	160	0	8.24	19.68	2.39	3.9	19.24
M22	Magurbeda	292.5	0	8.42	21.14	2.51	6.92	74.5
M16	Goria	163	0	4.87	13.23	2.72	6.28	57.07
M1	Patora	210	0	9.21	21.96	2.38	4.12	22.18
M14	Baragaon	200	0	8.63	24.21	2.81	4.61	25.81
M19	Ichhapur	172	0	8.08	19.07	2.36	4.02	20.95
M15	Takala	174.2	0	7.66	18.24	2.38	4.51	28.03
M18	Chhatikud	293	0	10.92	27.71	2.54	5.78	44.26
M17A	Burat	535.4	0	12.16	35.72	2.94	8.37	97.88
M17	Tulaghat	271	0	14.98	32.22	2.15	4.12	22.73
M20	Surubali	322.6	0	9.66	24.47	2.53	6.71	66.71
R5	Pipalpankha	130.2	0	7.39	14.58	1.97	3.93	22.18
Validation (2004 – 2005)								
KI8	Bhaskel	185.4	0	5.22	16.15	3.1	6.98	63.67
KI9	Kurumuli	304.6	0	12.6	33.43	2.65	5.54	38.4
M25	Sagada	89.4	0	3.55	10.58	2.98	5.04	30.82
M22	Magurbeda	132	0	6.45	15.5	2.4	4.58	26.08
M16	Goria	138	0	5.19	14.62	2.82	5.33	37.93
M1	Patora	281.6	0	8.79	24.04	2.74	6.35	58.45
M14	Baragaon	266	0	6.68	21.73	3.25	7.22	71.66
M19	Ichhapur	126	0	4.83	14.27	2.95	5.56	37.36
M15	Takala	132.8	0	8.1	18.73	2.31	3.22	11.57
M18	Chhatikud	267	0	8.56	23.03	2.69	6.24	56.17
M17A	Burat	342.2	0	9.77	27.64	2.83	7.33	73.78
M17	Tulaghat	169.2	0	11.91	25.32	2.13	3.45	13.53
M20	Surubali	172	0	10.36	19.76	1.91	3.57	18.39
R5	Pipalpankha	115	0	7.9	17.24	2.18	3.47	14.76

Table 2.8 Statistical parameters of discharge data (m^3/s) at Kantamal site

G&D site	Max.	Min.	Avg.	St. dev.	C_v	C_s	C_k
Calibration (2000 – 2003)							
Kantamal	12915.07	20.19	830.9	1523.35	1.83	4.78	28.32
Validation (2004 – 2005)							
Kantamal	11030.05	33.64	673.82	1196.79	1.78	5.43	35.6

Table 2.9 Statistical parameters of evaporation data (mm) at Dasapalla station in Kantamal sub-basin

Station name	Max.	Min.	Avg.	St. dev.	C_v	C_s	C_k
Calibration (2000 – 2003)							
Dasapalla	7.4	0	2.95	1.17	0.4	0.85	1.23
Validation (2004 – 2005)							
Dasapalla	9.7	0	3.12	1.61	0.52	1.16	1.96

3.1 BACKGROUND

This chapter presents a brief review of the measures taken for mitigation of floods in Orrisa and flooding problem of Mahanadi basin. A brief review of hydrological application of two soft computing methods namely Artificial Neural Network (ANN) and Fuzzy logic is also presented in this chapter.

Review of literature specifically on regional flood frequency analysis, L-moments, clustering methods, multi criteria decision methods and application of ANN and Fuzzy logic in flood forecasting is presented in corresponding chapters.

3.2 MEASURES TAKEN FOR MITIGATION OF FLOODS IN ORISSA

Orrisa state is a riverine state with eleven river basins. Several times the economy of the state has been crippled significantly by floods (Annexure–I). During the last five decades, a number of structural and non-structural measures have been taken to minimize the flood losses. As a part of structural measures, reservoirs namely Hirakud on the Mahanadi River, Rengali on the Brahmani River, Upper Kolab on Kolab River and Upper Indravati on Indravati River have been constructed. A total of 7138 kilometers of protective embankments and 1952 spurs have been constructed in different basins particularly in the deltaic areas to control the flood and salinity ingress (State Water Plan, 2007).

3.3 FLOODING PROBLEM OF MAHANADI

Mahanadi causes lot of devastation during flood time. During the period 1868 to 1946, the delta portion faced 63 floods with magnitude 1 million cusecs or more. The highest flood was observed in 1834 carrying a discharge of 1.571 million cusecs. Hence to mitigate the floods in deltaic portion, Hirakud dam was constructed in 1957. The embankments have also been constructed in the lower reach of Mahanahi. The Mahanadi embankment system in lower reach is safe for a flood discharge of 28300 m^3/s (1 million ft^3/s) (Khatua and Mahakul, 1999). Any discharge beyond 28300 m^3/s can create a flood like situation in delta. During the post-Hirakud period (1958 to 2008) a total of 19 floods have crossed the limit of 28300 m^3/s (Table 3.1). Out of these 19 events, 6 are due to Hirakud release and 13 are due to contribution of intercepted catchment.

Table 3.1 Flood peaks at Mundali and the corresponding contributions

Sl. No.	Year	Date	Peak flood at Mundali (m3/s)	Contribution towards peak at Mundali (m3/s)		Flood due to
				Hirakud dam release (D.R)	Intercepted catchment between Hirakud and Mundali (I.C)	
1	1958	19th July	33960.0	18593.1	15366.9	D.R
2	1959	14th Sept.	35516.5	14319.8	21196.7	I.C
3	1961	11th July	36365.5	30875.3	5490.2	D.R
4		18th July	32601.6	16866.8	15734.8	D.R
5		8th Sept.	33054.4	11518.1	21536.3	I.C
6		16th Sept.	36931.5	17602.6	19328.9	I.C
7	1969	1st Aug.	29941.4	2377.2	27564.2	I.C
8	1980	22nd Sept.	35601.4	33365.7	2235.7	D.R
9	1982	31st Aug.	44827.2	254.7	44572.5	I.C
10	1991	14th Aug.	35969.3	3282.8	32686.5	I.C
11	1992	30th July	32092.2	679.2	31413.0	I.C
12		21st Aug.	31413.0	2971.5	28441.5	I.C
13	1994	13th July	28979.2	17263.0	11716.2	D.R
14		6th Sept.	30592.3	10612.5	19979.8	I.C
15	2001	17th July	39620.0	12933.1	26686.9	I.C
16		20th July	39846.4	24139.9	15706.5	D.R
17	2003	30th Aug.	38205.0	2377.2	35827.8	I.C
18	2006	23rd Aug.	37016.4	1584.8	35431.6	I.C
19	2008	20th Sept.	44742.3	11065.3	33677.0	I.C

(*D.R= Dam Release, I.C = Contribution of Intercepted Catchment)

Under the present circumstances, construction of further flood control reservoirs is not possible due to large scale submergence and other environmental and ecological aspects. Hence, efforts have been made by DOWR in other directions to minimise the losses due to floods.

In 2007, a Flood Management Information System (FMIS) Cell was established in Bhubaneswar. This cell provides real time information on early flood warning, possible flood inundation and its impact by using Remote Sensing and Geographical Information System. Three automated sensors at Belagaon, Boudh and Naraj to provide instant water levels of Mahanadi have also been established. Since 2010, FMIS cell is providing catchment weighted rainfall and isohyetal maps on daily basis. From June 24, 2011 mobile phone SMS service has also been started for dissemination of flood information to all. Presently it is being operated by Mahanadi Barrage Division, Cuttack and providing the instant water levels at Naraj.

3.3.1 Existing Flood Forecasting System for Downstream Reach of Hirakud Dam

The catchment of Hirakud downstream part is divided into 5 parts, namely Hirakud to Khairmal, entire Tel catchment, Khairmal to Barmul, Barmul to Mundali and the deltaic part (Fig. 3.1). Each part is again identified as per the Thiessen area associated with the existing raingauges. A runoff coefficient is fixed for each Thiessen area which is updated daily keeping in view the previous day rainfall and current day flow observations. The product of runoff coefficient and daily rainfall gives discharge contributions for each segment. Travel time from one segment to further downstream is computed based on past flood data. Based on the daily rainfall and discharge released from Hirakud reservoir a tabular chart is prepared for deciding the discharge to reach Khairmal, Barmul and Mundali.

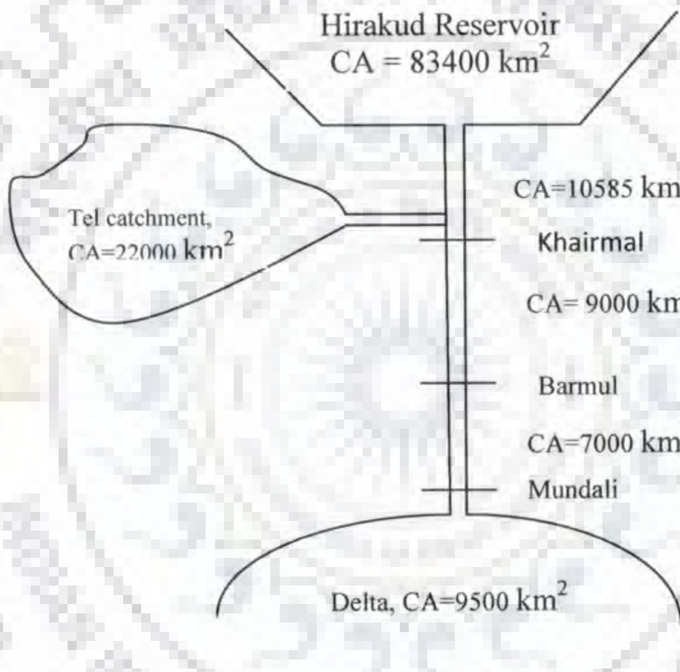


Fig. 3.1 Schematic downstream of Hirakud reservoir

The present forecasting system can be improved further by introducing the soft computing techniques based on ANN and Fuzzy logic. These techniques are briefly reviewed in the subsequent section.

3.4 SOFT COMPUTING TECHNIQUES

Soft computing refers to a collection of computational techniques in computer science, artificial intelligence, machine learning and some engineering disciplines which attempt to study, model and analyses very complex phenomena, for which conventional methods do not yield low cost, analytical and complete solution.

The guiding principle of soft computing as per Zadeh (1965) is "Exploit the tolerance for imprecision, uncertainty, partial truth and approximation to achieve tractability, robustness and low solution cost". The neural and genetic computing in soft computing came at a later stage in 1980. In the present study neural networks and fuzzy logic based methods have been used. These are briefly described in subsequent sections.

3.4.1 Neural Networks

ANNs have been developed as a generalization of mathematical models of neural biology and are based on following rules:

(i) Information processing occurs at many single elements called nodes, also referred to as units of neurons.

(ii) Signals are passed between nodes through connection links.

(iii) Each connection link has an associated weight that represents its connection strength

(iv) Each node typically applies a nonlinear transformation called activation function to its net input to determine its output signal.

ANNs are massively parallel systems composed of many processing elements connected by links of variable weights. The network consists of layers of neurons, with each layer being fully connected to the preceding layer by interconnection of strengths or weights (w). Figure 3.2 illustrates a three layer neural network consisting of input layer (L_i), hidden layer (L_H) and output layer (L_o) with the inter connection of weights W_{ih} and W_{ho} between layers of neurons. In multi-layered perception, hidden layer means second layer of processing elements or units in between the input and output layers that increases computational power. In principle, the hidden layer can be more than one layer. In practice, the number of neurons in this layer is evaluated by trial and error.

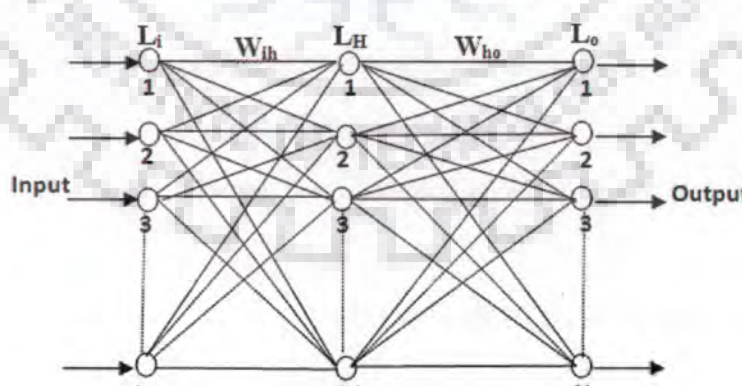


Fig. 3.2 Configuration of three-layer neural network

History of ANN stems from the work of Warren McCulloch and Walter Pitts who outlined a model of a simple neural network with electronic circuits in 1943.

The hierarchical background and review of application of artificial neural network in hydrology has been nicely documented by ASCE task committee in ASCE (2000 a, b). Further work in this area has been reviewed and documented in section 3.6.

3.4.2 Fuzzy Logic

Fuzzy logic is another area of soft computing that has been applied in different engineering fields. These concepts were introduced by Lotfi A. Zadeh (Zadeh, 1965). Fuzzy logic is a superset of conventional Boolean logic that has been extended to handle imprecise data and the concept of partial truth. A brief description of fuzzy logic theory and concept is given in the subsequent paragraph for the sake of completeness of the work and easy understanding of the readers.

In crisp logic, the values acquired by proposition are two valued namely True and False, which may be treated numerically equivalent to (0, 1). However, in fuzzy logic, true valued are multi-valued such as absolutely true, partly true, absolutely false, very true and so on are numerically equivalent to 0-1. The fuzzy set theory is also an effective tool to handle the problems of uncertainty.

In fuzzy logic, variables are fuzzified using different membership functions that define the membership degree to fuzzy sets. These variables are called linguistic variables. Fuzzy algorithms are formed by the union of fuzzy rules using operators. The important operators IF, THEN, AND, OR have significant impact on model performance. Fuzzy systems are defined by a number of fuzzy rules, a number of membership functions and mechanisms to apply logical operators. The outputs are defuzzified to get the crisp values (Lohani et al., 2007a).

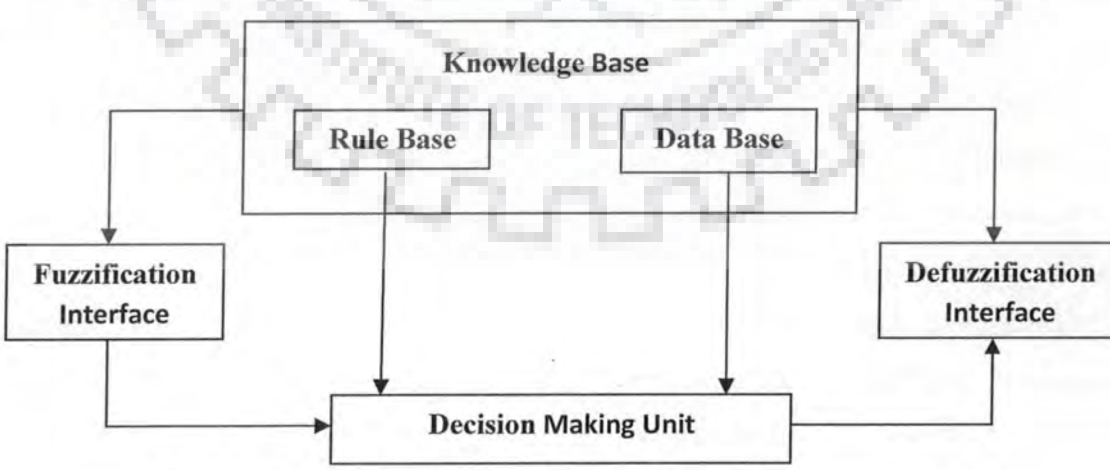


Fig. 3.3 Schematic representation of Fuzzy Inference System

There are numerous successful applications of fuzzy systems in control and modeling. They are suitable for situations where an exact model of a process is either impractical or very costly to build, but an imprecise model based on existing human expertise can do the job. In such situations, fuzzy systems are considered the best alternatives. Fuzzy sets are an aid in providing information in a more human comprehensible or natural form and can handle uncertainties at various levels. Fuzzy logic has made much progress since Zadeh (1965) initiated it. The further developments are as follows,

Zadeh (1973) introduced the concept of a linguistic variable, that is, a variable whose values are words rather than numbers. The concept of a linguistic variable has played and is continuing to play a pivotal role in the development of fuzzy logic and its applications.

Mamdani and Asssilian (1975) proposed a fuzzy inference system popularly known as Mamdani fuzzy inference system. The Mamdani fuzzy inference system was the first attempt to control a steam engine and boiler combination by a set of linguistic control rules obtained from experienced human operators. In Mamdani's application, two fuzzy inference systems were used as two controllers to generate the heat input to the boiler and throttle opening of the engine cylinder, respectively to regulate the stream pressure in the boiler and the speed of the engine. Since the plant takes only crisp values as inputs, therefore a defuzzifier was used to convert the fuzzy set into a crisp value.

Tsukamoto (1979) proposed a fuzzy model in which the consequent of each fuzzy if then rule is represented by a fuzzy set with a monotonical membership function. As a result the inferred output of each rule is defined as a crisp value induced by the rule's firing strength. Since each rule infers a crisp output, the Tsukamoto fuzzy model aggregate each rule's output by the method of weighted average and thus avoids the time consuming process of defuzzification. The Tsukamoto fuzzy model is not used often since it is not as transparent as either the Mamdani or Sugeno fuzzy models.

Pedrycz (1984) presented the identification algorithm in fuzzy relational model. Fuzzy relational models, which can be regarded as generalization of the linguistic model, encode associations between linguistic terms defined in the system's input and output domains by using fuzzy relations.

Takagi and Sugeno (1985) developed a model where the consequence is a crisp function of the antecedent variables rather than a fuzzy proposition. It can be seen as a combination of linguistic and mathematical regression modelling in the sense that the antecedents describe fuzzy regions in the input space in which the consequent functions are valid.

Jang (1993) proposed a class of adaptive networks that are functionally equivalent to fuzzy inference systems. The proposed architecture is referred to as ANFIS which stands for Adaptive Network based Fuzzy Inference System or semantically equivalently, Adaptive Neuro Fuzzy Inference System. The proposed scheme also described how to decompose the parameter set to facilitate the hybrid learning rule for ANFIS architectures representing both the Sugeno and Tsukamoto fuzzy models. It was also demonstrated that under certain minor constraints, the radial basis function network (RBFN) is functionally equivalent to the ANFIS architecture for the Sugeno fuzzy model.

Chiu (1994) presented an efficient method for estimating cluster centers of numerical data. This method can be used to determine the number of clusters and their initial values for initializing iterative optimization-based clustering algorithm. When combined with linear least squares estimation, it provides an extremely fast and accurate method for indentifying fuzzy models. Further, Mathworks (1994) introduced" The Fuzzy Logic Toolbox" for MATLAB as an add-on component to MATLAB.

Jang and Mizutani (1996) presented the results of applying the Levenberg-Marquardt method, a popular nonlinear least squares method, to the ANFIS architecture. Through empirical studies, they discussed the strength and weakness of using such an efficient nonlinear regression techniques for neuro-fuzzy modelling and explained the trade-offs between mapping precision and membership function interpretability.

Nauck and Kruse (1997) mention that neuro-fuzzy systems have recently gained a lot of interest in research and application. Neuro-fuzzy models are fuzzy systems that use local learning strategies to learn fuzzy sets and fuzzy rules. Neuro-fuzzy techniques have been developed to support the development of e.g. fuzzy controllers and fuzzy classifiers. The authors discuss a learning method for fuzzy classification rules.

Zadeh (1998) mentions that the status of fuzzy logic in 1998 is vastly different from what it was in 1978. He further states that the mathematical foundations of fuzzy logic are well established; the basic theory is in place; the impact of fuzzy logic on the basic sources and especially on mathematics, physics and chemistry is growing in visibility and importance; and fuzzy logic based applications are extending in a wide variety of directions. Dubois et al. (1998) state that the real power of fuzzy logic lies in its ability to combine modeling (constructing a function that accurately mimics the given data) and abstracting (articulating knowledge from the data).

Chen et al. (2000) proposed a new scheme to estimate the membership values for fuzzy set. The scheme takes input from empirical experimental data which reflects the expert's knowledge on the relative degree belonging to the members. First, they suggested an alternative (indirect) index for the expert(s) to submit. The index reflects the expert's assessment on the comparison of the degree belonging of each pair of elements. Second, based on the raw data which is generated via the use of the index, they proposed an optimization framework for calibration.

Shi and Mizumoto (2001) improved a neuro-fuzzy learning algorithm based on the fuzzy clustering method. In this approach, before learning fuzzy rules typical data were extracted from training data by using fuzzy c-means clustering algorithm, in order to remove redundant data and resolve conflicts in data, and make them as practical training data.

Abonyi et al. (2002) introduced a new clustering algorithm, that can be easily represented by an interpretable Takagi Sugeno fuzzy model. Similar to other fuzzy clustering algorithms, "modified Gath-Gevea algorithm" was employed in search of clusters.

Angelov (2004) developed and tested a recursive approach for adaptation of fuzzy rule-based model. Cluster centers calculated using on-line clustering of the input-output data with a recursively calculated spatial proximity measure are then used as prototypes of the centres of the fuzzy rules.

Rajsekharan and Pai (2004) mention that fuzzy set theory is an excellent mathematical tool to handle the uncertainty arising due to vagueness. Fuzziness means 'vagueness'. The fuzzy systems approximate functions. They are universal approximators if they use enough fuzzy rules. In this sense fuzzy systems can model any continuous function or system.

Aqil et al. (2006) mention that traditionally, the multiple linear regression technique has been one of the most widely used models in simulating hydrological time series. However, when the nonlinear phenomenon is significant, the multiple linear will fail to develop an appropriate predictive model. Recently, neuro-fuzzy systems have gained much popularity for calibrating the nonlinear relationships. The authors evaluated the potential of a neuro-fuzzy system as an alternative to the traditional statistical regression technique for the purpose of predicting flow from a local source in a river basin.

The applications of soft computing techniques in regional flood frequency analysis and flood forecasting are briefly reviewed in next section.

3.5 APPLICATION OF SOFT COMPUTING TECHNIQUES IN REGIONAL FLOOD FREQUENCY ANALYSIS AND FLOOD FORECASTING

The principal constituents of soft computing include fuzzy logic, neural network, and genetic algorithm. There are numerous applications of these constituents (both individually and combination of two or more) in the area of water resources and environmental systems. These range from development of data driven models to optimal control strategies to assist in more informed and intelligent decision making process. Availability of data is critical to such applications and having scarce data may lead to models that do not represent the response function over the entire domain. Applications of soft computing techniques in regional flood frequency analysis and flood forecasting are reviewed in subsequent sections.

3.5.1 Regional Flood Frequency Analysis

Delineation of hydrological homogeneous regions is one of the prime requirements of regional flood frequency analysis. Soft computing techniques are being used recently for delineating homogeneous regions. Earlier Thandeveswara and Sajjikumar (2000), Shi (2002), Bhatt (2003), Lim and Lye (2003), Jingyi and Hall (2004), Chavoshi and Soleiman (2009) and Dikbas et al. (2011) used soft computing techniques for delineations of homogeneous regions.

Thandeveswara and Sajjikumar (2000) have used ART-II technique in Artificial Neural Network in classification of river basin for finding homogeneous regions. They suggested that if C_v of C_v of a cluster is greater than 0.4 the region is highly heterogeneous. They suggested for overall objective of clustering as

i) to have statistically acceptable homogeneity and
ii) to have sufficient data in each cluster for further hydrologic studies.

Shi (2002) has applied clustering technique by using Fuzzy clustering technique and neural network.

Bhatt (2003) in his study has made critical evaluation of conventional techniques such as Ward's method and K-mean method and modern techniques like Fuzzy C-mean and Kohonen (ANN based) method. Lim and Lye (2003) used Hierarchical clustering (average linkage method) in order to delineate homogeneous sub-regions in Sarawak, Malaysia. They have also done appropriate scaling of catchment characteristics to ensure that these factors fell between zero and unity.

Jingyi and Hall (2004) have compared the result of K-mean, Fuzzy C-mean, Hierarchical clustering and Kohonen Self Organising Feature Map for getting the homogeneous region in Gan-Ming river basin of China. They have also utilized the suitability of Kohonen map for finding the number of clusters as well as sites allotted.

Chavoshi and Soleiman (2009) have applied conventional cluster analysis as well as Fuzzy Logic theory on regionalization of 70 catchments in north of Iran.

Regionalization along with clustering approach is also applied in low flow and rainfall analysis.

Dikbas et al. (2011) have recommended fuzzy C- mean (FCM) cluster method for classifying the precipitation series and for identifying the hydrologically homogeneous regions in a Turkish basin. They also focused on choice of appropriate cluster method and variables of the basin.

Besides those the soft computing techniques have also been applied by Shu and Burn (2004) to derive homogeneous region on similarity measure using Fuzzy Expert Systems (FES). Shu and Ouarda (2008) used Adaptive Neuro-Fuzzy Inference Systems (ANFIS) as a mechanism for identifying the hydrological regions by generating knowledge from hydrometric station network in southern Quebec. Ouarda and Shu (2009) introduced Artifical Neural Networks (ANNs) to obtain improved regional low-flow estimates at ungauged sites in the province of Quebec, Canada. Kumar (2009) applied ANN and Fuzzy logic for finding the growth factors in regional flood frequency analysis of sub-zones of India and compared with that obtained from L-moment analysis.

3.5.2 Flood Forecasting

Halff et al. (1993) applied a 3 layer NN model to portray the hydrographs recorded by the USGS at Bellvue, Washington. They used observed rainfall hyetographs as input. Since then several studies for rainfall-runoff modeling using NN models have been carried out.

Zealand et al. (1999) investigated the utility of ANN for short term forecasting of streamflow in a portion of the Winnipeg river system in North West Ontario, Canada of catchment 20000 sq km. using quarter monthly time interval. Explores the capability of ANNs and compares performance of this tool to conventional approaches used to forecast streamflow.

Coulibaly et al. (2000) used MLFFNN for real time reservoir inflow forecasting. The proposed method by the authors take advantage of both Levenberg- Marquardt BP and cross validation technique to avoid underfitting or overfitting and enhances generalization performance. Some of the recent applications of ANNs in hydrology include comparison of ANNs and empirical approaches for predicting runoff (Anmala et al. 2000)

Dawson and Wilby (2001) compared ANN methods with more conventional statistical models. Research on extraction of hydrological rules from ANN weights and on the development of standard performance measures that penalize unnecessary model complexity.

Rajurkar et al. (2002) used a linear multiple–input-single-output (MISO) model coupled with ANN for developing rainfall-runoff relationship for river Narmada, India. The model provided a systematic approach for runoff estimation.

Jain and Indurthy (2003) made comparative analysis of event based rainfall runoff modeling techniques, deterministic, statistical and artificial neural networks.

Xiong et al. (2004) recommended the ANN model with back propagation method which updates the weights of ANN at each time step according to the latest forecast error for use in real time flow forecasting. Lekkas et al. (2004) utilizes various types of ANNs in an attempt to assess the relative performance of existing models. Ali Efenti, a sub-catchment of the river Pinios (Greece), is examined and the results support the hypothesis that ANNs can produce qualitative forecasts. A 7-hour ahead forecast in particular proves to be of fairly high precision, especially when an error prediction technique is introduced to the ANN models. Nayak et al. (2004) applied ANFIS for time series modeling of river flow in Baitarani basin.

Wu et al. (2005) demonstrated an application of ANNs for watershed runoff and stream flow forecasts.

Teschl and Randeu (2006) applied statistical analysis to reduce amount of data and find an appropriate input vector for an ANN model using raingauge and weather radar data to predict the runoff of a small Alpine catchment in Austria. Bardosy (2006) has found the suitability of fuzzy logic as an alternative to statistical methods and to quantify uncertainties in modeling natural systems. Raghuwanshi et al. (2006) mention that accurate estimation of both runoff and sediment yield is required for proper watershed management. The ANN models were developed, to predict both runoff and sediment yield on a daily and weekly basis, for a small agricultural watershed.

Bhattacharya et al. (2007) developed a simulation methodology using a trained ANN model to approximate the three dimensional density dependent flow and transport processes in a coastal aquifer. Nor et al. (2007) applied RBF-ANN for rainfall-runoff modeling of a Malayasian catchment and predicted the streamflow hydrograph better than HEC-HMS model. Nayak et al. (2007) explored the potential of integrating two different artificial intelligence techniques, namely neural network and fuzzy logic, effectively to model the rainfall-runoff process from rainfall and runoff information. The integration is achieved through representing fuzzy system computations in a generic ANN architecture, which is functionally equivalent to a fuzzy inference system. The model is initialized by a hyper-ellipsoidal fuzzy clustering (HEC) procedure, which identifies suitable numbers of fuzzy if-then rules through proper partition of the input space.

Kisi (2008) demonstrated the application of different artificial neural network (ANN) techniques for the estimation of monthly streamflows. Different ANN techniques, namely, feed forward neural networks (FFNN), generalized regression neural networks (GRNN) and radial basis ANN (RBF) were used in one-month ahead streamflow forecasting and the results were evaluated. Based on the results, the GRNN was found to be better than the other ANN techniques in monthly flow forecasting. Firat and Gungors (2008) investigated the applicability of an ANFIS model to the forecasting of flow time series in a Turkish river.

Kisi (2009) proposed the application of neural network and wavelet conjuction model for flood forecasting of daily intermittent streamflow for Thrace region of Turkey. Hong and White (2009) introduced the dynamic neuro-fuzzy local modeling system (DNFLMS) that is based on a dynamic Takagi Sugeno (TS) type fuzzy inference system with on-line and local learning algorithm for complex dynamic hydrological modeling tasks. The proposed DNFLMS was applied to develop a model to forecast the flow of Waikoropupu Springs, located in the Takaka Valley, South Island, New Zealand, and the influence of the operation of the 32 Megawatt Cobb hydropower station on spring flow.

Nandalal and Ratanayake (2011) tested that how effectively the risk with respect to flood can be assessed using fuzzy approach on the study area Kalu-Ganga river basin of Srilanka. They have taken flood extent and mean flood depth as hazard indicators while density and dependency ratio as vulnerability indicators. A methodology was proposed and applied to asses risk assuming the above indicators as fuzzy variables.

Kar et al. (2010) has applied ANN on peak floods at lower reach of Mahandi basin for predicting floods at Mundali. The results are compared with statistical methods and ANN results are found suitable and efficient.

Lohani et al. (2005a, b) developed a real time fuzzy based model at Mandla site of Narmada river using discharge and rainfall data. Lohani et al. (2006, 2007a, b) developed a fuzzy based stage–discharge relationship for Jamatara gauging site. Lohani et al. (2011) compared the performance of ANN, fuzzy logic and linear transfer function for rainfall–runoff modeling. They have also investigated the potential of Takagi Sugeno fuzzy model and impact of soil moisture condition in the performance of the daily rainfall-runoff model.

Mukerji et al. (2009) have applied ANN, Adaptive Neuro Fuzzy Inference System (ANFIS) and Adaptive Neuro-GA Integrated System (ANGIS) model for flood forecasting in Ajay river basin. The ANGIS and ANFIS models have given a better performance over ANN models. Tiwari and Chatterjee (2010) investigated the uncertainty associated with hourly flood forecasting using bootstrap based ANN and found these models with confidence bounds can improve their reliability for flood forecasting.

Apart from frequency analysis and flood forecasting the soft computing techniques have been applied in other fields of hydrology also e.g. Awchi (2008) for prediction of reference crop evapotranspiration, Kisi (2010) has investigated the application of fuzzy genetic approach in modeling of reference evapotranspiration (ET0), Haghizadeh et al. (2010) for prediction of sediment yield in Sorkhab river and Mayilvaganan and Naidu (2011) for prediction of ground water levels.

3.6 CONCLUSIONS

The review of literature clearly indicates that the flood related problems of Mahanadi basin require high attention. The non-structural measures like flood frequency prediction and flood forecasting are the two best alternatives in the present scenario. There is strong need to improve the existing regional flood formulae in Mahanadi basin. Also the existing flood forecasting system for lower reach of Mahanadi basin needs improvement by introducing the soft computing based methods.

The details of the studies undertaken for the same are presented in the subsequent chapters.

CHAPTER 4 - REGIONAL FLOOD FREQUENCY ANALYSIS OF MAHANADI BASIN USING PRIORITISED VARIABLES

4.1 BACKGROUND

Selection of suitable site characteristics and number of clusters play an important role for finding homogeneous regions in regional flood frequency analysis. Present study investigates the partition of Mahanadi basin into homogeneous regions by applying different clustering techniques using less but influential variables. As such the entire basin is not hydro-meteorologically homogeneous. Principal component analysis (PCA) has been initiated in finding appropriate site characteristics (variables) as per priority. Out of seven variables four variables are selected on priority. Possible numbers of cluster are found by applying Kohonen self organization map and Andrews plot (AP). Other clustering techniques like Hierarchical clustering (HC), Fuzzy C-mean (FCM) and K-mean (KMean) are applied on prioritised variables to verify the result of clustering. The inter-comparison of clustering techniques gives the optimum number of sites to be placed in a particular cluster. The sites clustered as per FCM give better result as far as homogeneity is concerned. The entire basin is divided into two homogeneous clusters. The regional L-moment algorithm is used to test the homogeneity and to identify an appropriate underlying frequency distribution. An index flood is also stated relating to catchment characteristics using multiple linear regression approach. The results are compared with the earlier studies of flood frequency on this basin. The study faces the limitation of lesser data availability in order to predict longer return period values (Q_T).

Research findings of this chapter have been accepted by ASCE, Journal of Hydrologic Engineering entitled as "*Application of clustering techniques using prioritized variables in regional flood frequency analysis- a case study of Mahanadi basin, India*" on April 15, 2011 with doi:10.1061/(ASCE)HE.1943-5584.0000417. A brief review of literature has been provided here besides Chapter-3 and then methodology and overview of the data used in the study are provided. This is followed by interpretation of results and conclusions.

4.2 INTRODUCTION

Regional flood frequency analysis typically begins with a "region" that comprises a group of sites from which extreme flow events can be combined for improving the estimation of extreme flows at any site in the region. The requirements that a region should posses to ensure effective information transfer (Burn and Goel, 2000) are:

(i) The region should be hydrological homogeneous. This requirement arises from the need to ensure that extreme flow that is transferred to the target site is similar to the extreme flow information of that site.

(ii) The region should be identifiable which implies that a regional home can be readily determined for a new catchment, which may be un-gauged.

(iii)The region should be sufficiently large. Larger regions imply that more extreme flow information is incorporated into estimation of extreme flow quantiles.

As the size of the region is to be increased, there is a tendency for homogeneity of a region to be decreased. So there is a trade-off between first and third requirement. Application of L-moment along with cluster analysis is best utilized for delineating homogeneous regions.

L-moment approach: The estimation of a flood magnitude associated with a given return period is a crucial task for designing variety of engineering works and hydraulic structures. The research on flood frequency analysis has taken up with a varying intensity over a couple of decades. During seventies and eighties much effort was directed on developing efficient at site flood frequency procedure. Research on nineties was largely dominated by L-moments. The L-moments of a random variable were first introduced by Hosking (1986).They are analogous to conventional moments, but are estimated as linear combination of order statistics. The L-moments are defined as linear combinations of probability weighted moments (Hosking, 1986, 1990). In a wide range of hydrologic applications, L-moments provide simple and reasonably efficient estimators of characteristics of hydrologic data and of distributions parameters (Stedinger et al., 1992). Hosking and Wallis (1997) state that L-moments are an alternative system of describing the shapes of probability distribution. They are robust to outliers and virtually unbiased for small samples, making them suitable for flood frequency analysis including identification of distribution and parameter estimation. Applications of L-moment are largely found in the field of hydrology. The work of Kumar et al. (1999, 2003), Kumar and Chatterjee (2005), Borujeni et al. (2009), Sarkar et al. (2010) are some of the noteworthy works in this regard.

Clustering approach: The search for homogeneous region using cluster analysis earlier was done by DeCoursey (1973), DeCoursey and Deal (1974), Mosley (1981), Tasker (1982), Acreman and Sinclair (1986), Burn (1989), Nathan and McMohan (1990), Lim and Lye (2003). Jingyi and Hall (2004) had compared the result of K-mean, Fuzzy C-mean, Hierarchical clustering and Kohonen Self Organising Feature Map for getting the homogeneous region in Gan-Ming river basin of China. They have also utilized the suitability of Kohonen map for finding the number of clusters as well as sites allotted. Chavoshi and Soleiman (2009) applied conventional cluster analysis as well as Fuzzy Logic theory on regionalisation of 70 catchments in north of Iran. Regionalisation along with clustering approach is also applied in low flow and rainfall analysis. Fovell and Fovell (1993) used hierarchical clustering (HC) in combination with Principal component analysis (PCA) for identifying climatic regions of the United States based on monthly rainfall and temperature data. For obtaining extreme rainfall values L-moment approach has been applied by Eslamian and Feizi (2007) in Zayandehrood basin of Iran. Eslamian and Biabanaki (F2008) have applied cluster analysis and Andrews plot in Kharkeh basin for regionalization of low flow. Stambuk et al. (2007) have investigated possible application of the Kohonen Self Organizing Maps (SOM) to the social sciences data clustering, and compare the results of the procedure to the Principal Component Analysis (PCA) and Hierarchical clustering methods. Beaulieu et al. (2009) has compared two new Bayesian change point techniques with eight other techniques to detect in-homogeneities in climatic series.

4.3 STUDY AREA

The study area Mahanadi basin is a major east flowing river in peninsular river system of India. It originates near Pharasiya village of Raipur district of Chhatisgarh state. The total drainage area of river basin is 141589 sq. km. The basin is en-compassed within the geographical co-ordinate of 80^0-30' to 86^0-50' of East Longitude and 19^0-20' to 23^0-35' of North Latitude. The basin is physically bounded by Central Hills in the North, Eastern Ghat in the South, Maikala hill range in the West and by Bay of Bengal in the East. The basin is largely divided into four parts such as Central table land, Northern plateau, Eastern ghats and Coastal plain. A lot of geographic and climatic variation is seen in the basin due to its large size. A large reservoir named Hirakud is situated at center of the catchment draining 83000 sq km. into it. The study area is shown in Fig.4.1 mentioning the G&D (gauge and discharge) sites and location of major reservoir Hirakud. The flood frequency analysis is not done in the regulated part of the catchment.

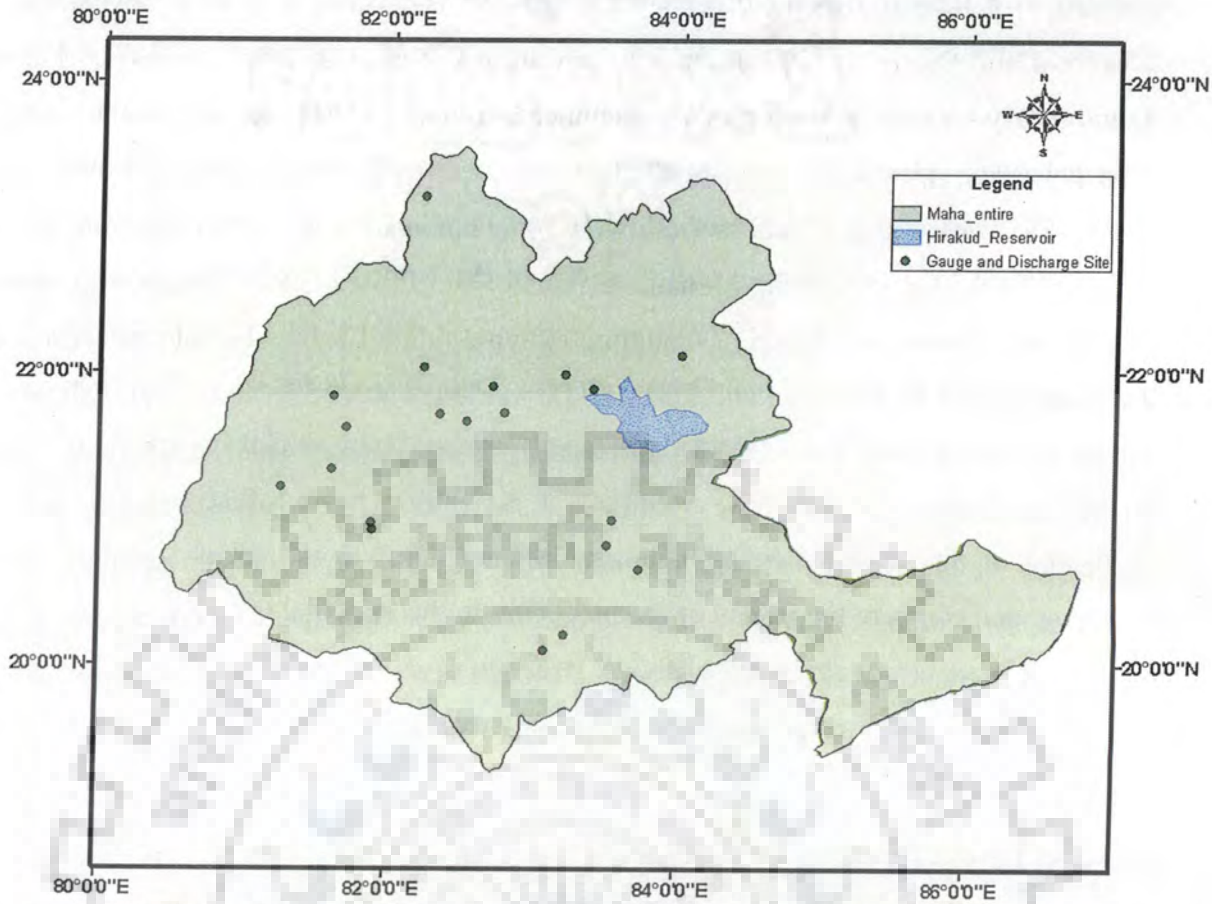


Fig. 4.1 Study area with gauge and discharge site of Mahanadi basin

4.4 DATA AVAILABILITY

In the present study, annual maximum flood series (AFS) data of 20 gauging stations (Table 4.1) spread over entire basin are collected. The study area with gauge and discharge sites has been shown in Figure 4.1. The details like station identification number (Station Id), station name, station years, latitude, longitude, maximum observed discharge at the station (Q_{max}), normal rainfall (NR) and maximum one day rainfall (1D) are collected from state government office. The catchment descriptors like catchment area (CA), station elevation above mean sea level (SH), average slope of watershed (SL), drainage density (DD) and longest stream length (RL) are derived from digital elevation model (DEM) of study area. Either station id or station names are used in further applications.

Table 4.1 Location of gauge and discharge (G and D) stations with their Id number

Station Id	Station name	Latitude (deg-min-sec)	Longitude (deg-min-sec)
1	Sundergarh	22-06-55	84-00-40
2	Kurubhatta	21-59-15	83-12-15
3	Ghatora	22-02-04	82-13-34
4	Jondhra	21-43-00	82-20-34
5	Basantapur	21-43-18	82-47-27
6	Andhiyarkore	21-47-00	81-36-30
7	Bamanidhi	21-53-55	82-32-37
8	Rampur	21-39-00	82-31-00
9	Salebhatta	20-59-00	82-42-29
10	Baronda	20-55-06	81-52-56
11	Rajim	20-58-00	81-52-30
12	Kotni	21-13-02	81-14-19
13	Simga	21-37-33	81-41-36
14	Kantamal	20-38-49	83-43-55
15	Kesinga	20-11-51	83-13-30
16	Kelo	21-53-19	83-24-10
17	Mahendragarh	23-12-10	82-12-54
18	Pandigaon	20-05-35	83-05-00
19	Pathardihi	21-20-28	81-35-48
20	Sukuma	20-48-30	84-30-00

4.5 METHODOLOGY

The methodology consisted of following major steps:

(a) Checking the homogeneity of the entire basin.

(b) Selection of prioritised variables.

(c) Application of clustering methods.

(d) Selection of regional distribution.

4.5.1 Checking Homogeneity of the Entire Basin

4.5.1.1 Statistical measures of L-moment

Three statistical measures discordancy measure, heterogeneity measure and goodness of fit measure as per L-moment approach are used in regional studies. These measures are explained by Hosking and Wallis (1997). Here only the heterogeneity measure is discussed.

4.5.1.2 Heterogeneity measures

It is used to estimate the degree of heterogeneity and to assess whether they might reasonably be treated as homogeneous. Specifically, the heterogeneity measure compares between site variations in sample L-moments for the group of sites with that expected for a group of region. Hosking's Heterogeneity test fits 4 parameter Kappa distributions. A series of 500 simulations (N_{sim}) done and L-statistics of actual region is compared with a simulated series. The H-statistics are defined as,

$$H_i = (V_i - \mu_v) / \sigma_v \tag{4.1}$$

For each simulated region, the measures of variability V_i (where V_i is any of three measures V_1, V_2 and V_3) are calculated. From the simulated data the mean μ_v and standard deviation σ_v of the N_{sim} values of V_i are determined.

The critical H statistics for a region to be homogeneous is as mentioned below

$H<1$	homogeneous	(4.2)
$1 \le H \le 2$	possibly heterogeneous	(4.3)
$H>2$	definitely heterogeneous	(4.4)

Hosking and Wallis (1991) observed that statistics H_2 and H_3 based on measure of V_2 and V_3 lack the power to discriminate between homogeneous and heterogeneous regions but H_1 based on V_1 has much better discriminating power. So H_1 is treated as a much better indicator of heterogeneity measure. Also, H_1 was found to be a better indicator of heterogeneity in large regions, but has a tendency to give false indication of homogeneity for small regions (Rao and Hamed, 2000).

The measure H_2 indicates whether at-site and regional estimates will be close to each other. A large value of H_2 indicates whether or not the at-site and regional estimates will be in agreement, whereas a large value of H_3 indicates a large deviation between at-site estimates and observed data.

4.5.2 Selection of Prioritised Variables

4.5.2.1 Deriving the variables

For application of clustering methods each site should be presented with its individual characteristics (variables). Some common attributes under physical, hydrological and meteorological categories given by Parida (2004) are listed in Table-4.2.

Table 4.2 Some commonly used variables

Sl. No.	Variable category	Name of the Variables
1	Physical	Basin area, Average slope of basin, Elevation of gauging site, Length of main channel, General soil characteristics of the basin.
2	Hydrological	Annual average flows, Coefficient of variation and Coefficient of skewness of annual peak floods
3	Meteorological	Annual average rainfall, Other critical rainfall values (say 50 year-3 hour) rainfall or some such identified characteristics.

The variables should be selected in such a way that it should be more common and easily available to the practicing hydrologists and field engineers as well as it should represent both the statistical and physical importance towards runoff generation. The hydro-meteorological data used in this study are 1D and NR of the corresponding site. For obtaining physical variables DEM of the study area has been downloaded from freely available 90m SRTM data. The desired DEM is processed using the ARCGIS. The CA, SL, DD, RL, SH are derived for individual catchment using ARCGIS toolbox.

4.5.2.2 Discussion on variable selection

As the variables collected have different measuring units, are to be normalized between zero and unity. In order to determine which variables are more predominant, PCA has been applied initially. Number of principal components and its corresponding variances to be achieved depends upon the nature of study. Any PC having variance less than 1 contains less information than one of the original variables so can be dropped. The rule is regarded as Kaiser's rule (Kaiser, 1960) and retains those PCs whose variances exceed 1. The variables influencing these PCs are to be recorded as prioritized variable. However while prioritizing both statistical as well as physical importance of the variable should also be considered.

The statistical importance of the variables will come from the loadings of the PCs. The physical importance of individual variables is also discussed here. The CA as a variable speaks the size of the site and the smaller the size the chances are there it will have less contribution of runoff. SL is an important characteristic of a catchment as it gives an indication of the kinetic energy available for water to move towards the basin outlet, and it has been found to be related to total runoff and base flows (Bullock, 1988; Vogel and Kroll, 1992). DD is derived by dividing the sum of total stream length within a catchment by the catchment area, and is regarded as an important landscape characteristic. It is a measure of how dissected a basin is, and it is expected that DD affects the transformation of rainfall into runoff (Pitlick, 1994; Berger and Entekhabi, 2001). Fast flow will occur with area of high DD and steep SL. The RL represents the shape of the catchment for generating runoff. The 1D remains responsible for formation of flood peaks. A higher 1D can generate a peaky runoff rather than an average rainfall spread over longer time duration. The NR of a watershed speaks the usual rainfall receiving characteristics. A high NR getting watershed if receives a heavy 1D can generate a high flood. Usually the high normal rainfall areas get the maximum daily rainfalls. However SH has very less influence on runoff generation.

Hall and Minns (1999) applied fuzzy c mean algorithm for regionalisation by applying in two regions identified by United Kingdom Flood Studies Report (NERC, 1975). They have considered catchment area, main stream length, main stream slope, mean annual rainfall and soil index as features for analysis. Malekinezhad et al. (2010) in their study on Namak lake basin of Iran have taken 4 independent variables like water way length, mean annual precipitation, compactness (Gravelious) coefficient and mean annual temperature out of 14 independent variables.

4.5.2.3 Selecting suitable number of variable out of prioritized list

When the number of variable selected is large, a small but effective group of variable is to be considered for further analysis instead of using all variables. In this regard multiple regressions can be more appropriate towards choosing exact number of variables. The independent variables should again be selected keeping in view the underlying physical process as well as a good correlation with the dependent variable. It will be better if number of variable selected is same or somewhat more than the number of PCs with a reasonable correlation value. In the process relatively small loss of information may happen after dropping some of the less influencing variables.

4.5.3 Clustering Techniques

4.5.3.1 Application of clustering methods

The overall objectives of clustering (Thandeveswara and Sajikumar, 2000) are

(1) to have statistically acceptable homogeneity, and

(2) to have sufficient data in each cluster for further hydrologic studies.

In this study the number of clusters to which the basin should be optimally divided is decided on following factors,

- the result of Kohonen Self organizing map and Andrews plot, supported by other clustering methods.
- there should be at least seven sites in a group (Robson and Reed,1999)
- total station years in a pooling group as per 5T guideline in order to get a reasonable return period.

The performance of different clustering techniques has been evaluated on the basis of L-moments based statistical measures (Hosking and Wallis, 1993, 1997).

In application of clustering methods, it is first to decide what are the variables to be incorporated. In this study 7 variables are collected / calculated for each of the stations. But for clustering purpose only the top prioritized variables are considered instead of all.

. For identification of homogeneous regions from clustering techniques namely Kohonen Self organization map, Hierarchical clustering (Ward's method), K-mean, Fuzzy C-mean and Andrews plot methods have been used. The discussions on these techniques are given in Annexure-II.

4.5.3.2 Selection of Regional Distribution

The regional distribution has been adjudged on the basis of Z-statistics as follows:

Z-statistics

It indicates suitability of a candidate distribution to a data series and is appropriate for evaluating and comparing a distribution. The Z -statistics for the goodness of fit measure as defined by Hosking is

$$Z^{DIST} = (Z_4^{DIST} - Z_4 + B_4)/\sigma_4 \tag{4.5}$$

$DIST$ = a particular distribution, Z_4^{DIST} = L-kurtosis for fitted distribution, Z_4 = pooled L-kurtosis, B_4 = bias correction, σ_4 = estimate of sample variability of L-kurtosis. The Z^{DIST} value should be close to zero. However a value between -1.64 and 1.64 is considered to be suitable for a fitting distribution at 10 % significance level. While a number of distributions may qualify the goodness-of-fit criteria, the most potential will be one that has minimum $|Z^{DIST}|$ value.

4.6 RESULTS AND DISCUSSIONS

4.6.1 Heterogeneity Measure

By applying the L-moment algorithm of Hosking and Wallis the Heterogeneity measures i.e. H_1, H_2 and H_3 for the entire Mahanadi basin are 2.70, 2.12 and 2.05 respectively. Generalised Extreme Value (GEV), Generalized Normal (GN) and Person Type-III (PT-

III) are the derived suitable distributions. However it clearly indicates that all heterogeneity measures such as H_1, H_2 and H_3 indicate the basin as heterogeneous as the values obtained are above 2. Hence none of the distribution can be taken as granted. So now the task remains to form homogeneous sub-regions out of non-homogeneous region. Different techniques applied step by step are discussed below.

4.6.2 Cluster Formation

4.6.2.1 Standardization of variables

Before any application of the clustering methods the catchment characteristics (variables) is normalized (details in Annexure II). In our study 7 variables like catchment area (CA), station elevation (SH), normal rainfall (NR), maximum 1-day precipitation (1D), longest stream length (RL), average slope of each catchment (SL), drainage density (DD) of individual sites are used.

4.6.2.2 Selection of optimum variables

By applying PCA to 7 variables we get the variances with respect to Principal Components (PC) as mentioned in Table 4.3. Selecting only top 3 PCs with eigen values more than 1, which are giving nearly 75 percentage of the variance explained, there loadings are considered for finding the influence of the variable. The loadings of individual PCs are shown in Fig. 4.2 (a) to 4.2 (c) and their values are recorded in Table 4.4.

Table 4.3 Result of Principal Component Analysis

Principal Components	Eigen Value	Variance (percentage)	Cumulative variance
1	2.17	30.93	30.93
2	1.73	24.70	55.63
3	1.35	19.34	74.97
4	0.74	10.57	85.54
5	0.45	6.47	92.01
6	0.35	4.96	96.97
7	0.21	3.03	100.00

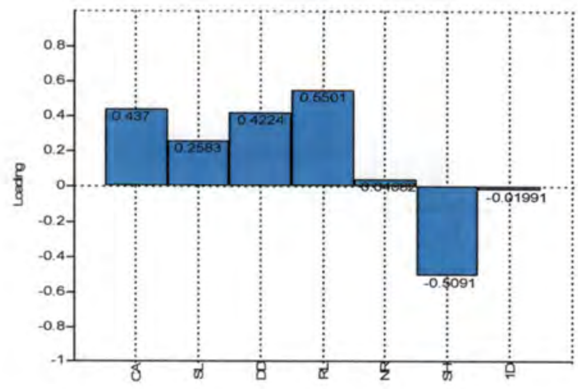


Fig. 4.2(a) Loading of different variables of PC-1

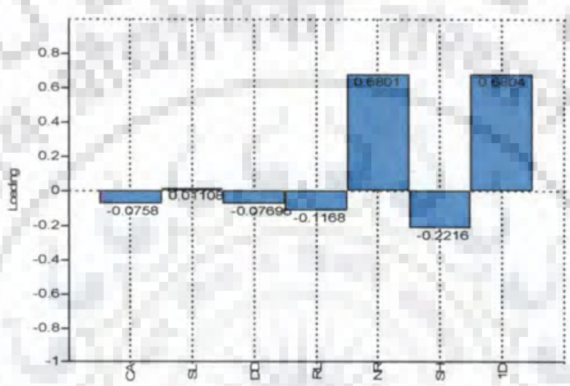


Fig. 4.2(b) Loading of different variables of PC-2

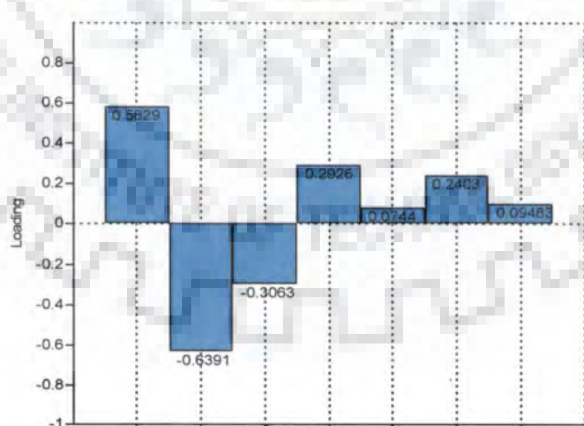


Fig. 4.2(c) Loading of different variables of PC-3

Table 4.4 Loading values of first 3 PCs

Variables	PC 1	PC 2	PC 3
CA	0.437	-0.0758	0.5829
SL	0.2583	-0.01108	-0.6391
DD	0.4224	-0.07696	-0.3063
RL	0.5501	-0.1168	0.2926
NR	0.0408	0.6801	0.0744
SH	-0.5091	-0.2216	0.2403
1D	-0.01991	0.6804	0.09483

From the selected 3 PCs the influential variables are chosen keeping both statistical and physical importance to the catchment. From Table 4.4 it is revealed that RL, SH, CA and dominate the loading of PC1, 1D and NR dominate the PC2 and PC3 by SL and CA. As per the loading values 1D, NR, SL, CA are the top 4 variables according to their individual loadings. Between RL and DD, the physical importance of DD and RL are very close in nature. Rather DD will be more representative towards representing the characters of a watershed. Again DD represents both area and river network whereas RL confined to longest stream. Similarly DD is also important than SH as latter has very little influence in runoff generation. So the final sequence as per importance of variables is as such 1D, NR, SL, CA, DD, RL and SH.

In order to reduce the number of variables from prioritized list and to create an effective variable list, the independent variables are put into multiple regressions with dependent variable (average Q_{max}). On regressing with 7 variables the R^2 value comes as 0.906 and dropping one inferior variable gradually from prioritized list the R^2 is reduced. By using 4 variables (ID, NR, SL, CA) the R^2 comes as 0.872 and using 5 variables (ID, NR, SL, CA, DD) the R^2 comes as 0.892. The variable CA is also a major indicator of R^2. As the numbers of PCs selected are 3 the number of possible variables selected may be somewhat more than 3. Even using 4 variables the information lost is very less. So for further analysis only 4 variables (ID, NR, SL, CA) are used instead of 7.

Existence of possible number of clusters: After getting the prioritized variables investigation is started for number of clusters. SOM is applied starting from all 7 variables and then reducing one by one. The same exercise is also done using AP. It is found that in all cases of SOM existence of two clusters are verified. Whereas AP does not give a recognizable plot with more variable. Stations which have similar properties would produce a band of similarly shaped curves. If a curve falls outside some margin, the given site can be assigned to a different group. The result of AP plots showing that there exists one uniform group whereas others are deflected from that group. The deflected parts are arranged into one group. However using 4 or 5 prioritised variable it has shown a good plot to identify the corresponding sites (Fig. 4.3).

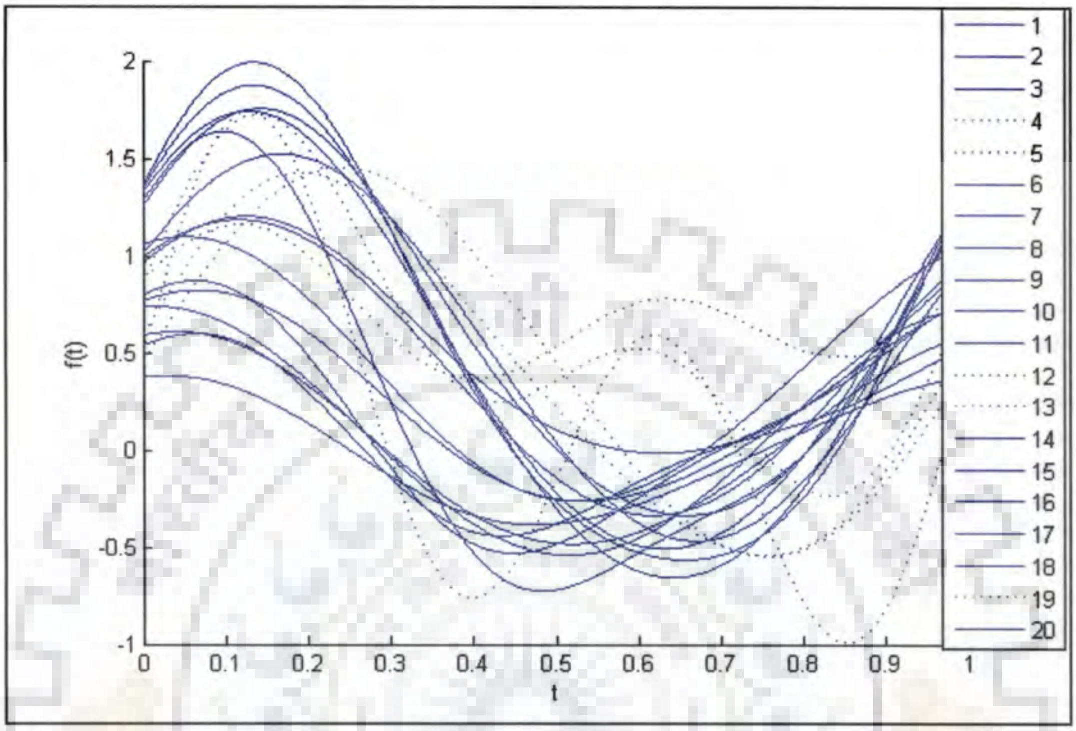


Fig. 4.3 Andrews plot with 4 prioritised variables for all 20 stations

The possible number of clusters and site allotments as Kohonen self organizing map plot is shown in Fig.4.4.

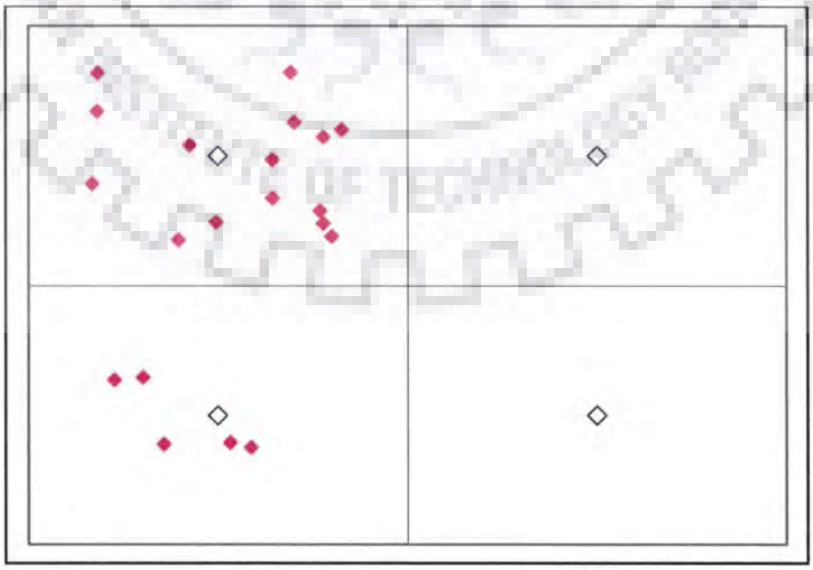


Fig. 4.4 Kohonen SOM plot with 4 prioritised variables for all 20 stations

This has been done using NNClust software. From SOM and deflection of Andrews plot maximum 2 clusters have been decided. Trying for more number of clusters for finding homogeneous region would have been resulted in drawbacks like less number of sites in a cluster, less station years and formation of small clusters with little identity. For verification two clusters are tried with 4 to 7 variables using all the clustering methods discussed.

4.6.2.3 Results of different clustering methods with heterogeneity measure

All discussed 5 clustering methods are applied on 4 to 7 variables. The corresponding results are noted in Table 4.5. It is observed from Table 4.5 that the sites allotted to each cluster and its corresponding heterogeneity measures. The HC result does not remain consistent with change in variables. With application of 4, 5 and 7 variables only H_2 results of cluster-1 remain as 0.75, 0.8 and 0.75 i.e. within homogeneous range. But other values of cluster 1 and 2 remain in possibly heterogeneous range. The dendrogram made with 4 variables is shown in Fig. 4.5.

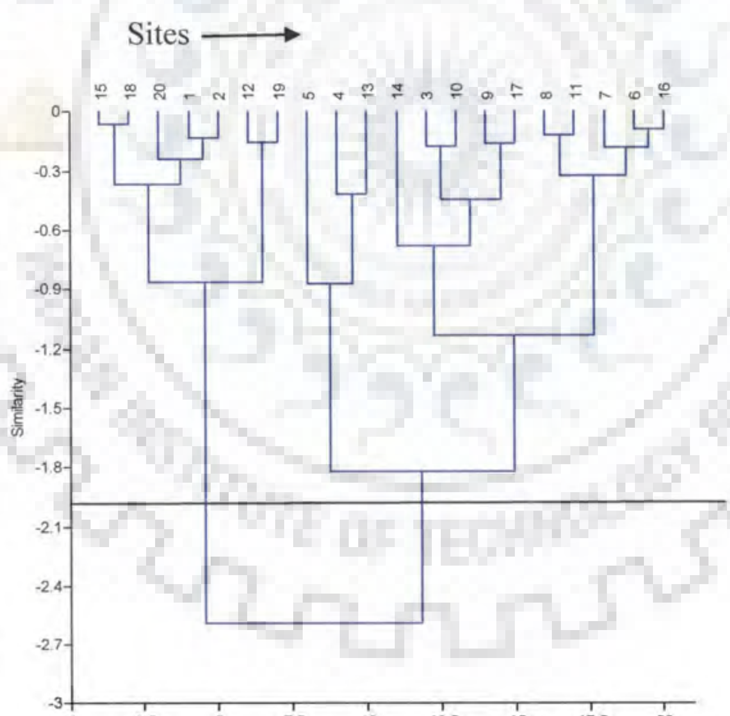


Fig. 4.5 Result of Hierarchical clustering using 4 variables (Dendrogram)

Table 4.5 Results of different clustering methods with heterogeneity measure as per prioritized variables (4-7)

Clustering method	Variable	4	5	6	7
HC	Cluster-1	3,4,5,6,7,8,9, 10,11,13,14, 16,17	3,4,5,6,7,8,9, 10,11,13, 14,15,16,17	3,5,6,7,8,10, 11,14,16	3,5,6,7,8,10, 11,14,16,17
	H_1,H_2,H_3	1.93,0.75,1.35	1.62,0.8,1.38	1.82,1.43,2.7	1.93,0.75,1.35
	Cluster-2	1,2,12,15,18, 19,20	1,2,12,18, 19,20	1,2,4,9,12,13,15, 17,18,19,20	1,2,4,9,12,13, 15,18,19,20
	H_1,H_2,H_3	1.09,2.18,1.01	1.14,2.48,1.4	0.24,0.5,0.23	1.18,2.01,0.96
KMean	Cluster-1	3,5,6,7,8,10, 11,13,14,16	3,5,6,7,8,10, 11,14,16	3,5,6,7,8,10, 11,14,16	3,5,6,7,8,10, 11,13,16,17
	H_1,H_2,H_3	1.69,1.45,2.4	1.82,1.43,2.1	1.82,1.43,2.1	1.85,0.13,0.64
	Cluster-2	1,2,4,9,12,15,17, 18,19,20	1,2,4,9,12, 13,15,17,18, 19,20	1,2,4,9,12,13, 15,17,18,19,20	1,2,4,9,12,14, 15,18,19,20
	H_1,H_2,H_3	0.47,0.86,0.12	0.24,0.5,0.23	0.24,0.5,0.23	0.8,2.12,1.86
FCM	Cluster-1	3,5,6,7,8,10, 11,14,16	3,5,6,7,8,10, 11,14,16	3,5,6,7,8,10, 11,13,14,16	3,5,6,7,8,10, 11,13,14,16
	H_1,H_2,H_3	1.82,1.43,2.1	1.82,1.43,2.1	1.69,1.45,2.4	1.69,1.45,2.4
	Cluster-2	1,2,4,9,12,13,15, 17,18,19,20	1,2,4,9,12, 13,15,17,18, 19,20	1,2,4,9,12,15, 17,18,19,20	1,2,4,9,12,15,1 7,18,19,20
	H_1,H_2,H_3	0.24,0.5,0.23	0.24,0.5,0.23	0.47,0.86,0.12	0.47,0.86,0.12
SOM	Cluster-1	6,7,8,11,16	7,11,13,16	5,7,9,12,13	3,4,5,6,7,8,9, 10,11,13,14, 16,17
	H_1,H_2,H_3	0.65,0.2,2.02	1.94,0.44,1.4	0.02,-0.6,-0.62	2.01,0.8,1.35
	Cluster-2	1,2,3,4,5,9,10, 12,13,14,15,17, 18,19,20	1,2,3,4,5,6, 8,9,10,12, 14,15,17,18, 19,20	1,2,3,4,6,8,10, 11,14,15,16,17, 18,19,20	1,2,12,15,18, 19,20
	H_1,H_2,H_3	0.74,1.15,0.61	0.94,1.6,1.05	2.77,2.94,2.82	1.09,2.18,1.01
AP	Cluster-1	4,5,12,13,19	5,7,9,12,13	Not well distinguished	Not well distinguished
	H_1,H_2,H_3	-0.43,-0.96,-1.31	0.2,-0.6,-0.62		
	Cluster-2	1,2,3,6,7,8,9,10,1 1,14,15,16,17,18, 20	1,2,3,4,6,8,10 ,11,14,15,16, 17,18,19,20	Not well distinguished	Not well distinguished
	H_1,H_2,H_3	2.0,1.87,2.34	2.94,3.0,2.87		

Applying KMean in 4 to 6 variables in cluster1 the result of H_1 and H_2 remains in possibly heterogeneous range while H_3 shows heterogeneity. With application of 7 variables H_1 value remains in possible heterogeneous range while other values remain homogeneous. Almost all cluster-2 values are in homogeneous range except H_3 value when 7 variables are applied.

The FCM result is quite uniform. It shows only heterogeneity for H_3 of cluster-1 under all variables but the cluster-2 remains complete homogeneous.

SOM shows somewhat acceptable homogeneity with 4 and 5 variables. But the number of sites allotted to cluster-1 is very less so less station year will be accommodated, which will make difficult to predict for long return period. So existence of less sites makes the cluster-1 meaningless.

Again AP shows homogeneity for cluster-1 with 4 and 5 variables but the result of cluster-2 is not acceptable on homogeneity ground. The sites allotted to cluster-1 is again very less. The AP plot remains undistinguishable in case of 6 and 7 variables, so not considered for the analysis here. It also found that some sites remain very common to at least more than one clustering methods under different variables. The sites like 6, 7, 8, 10, 11, 14 and16 remains common sites of cluster-1. Like 1, 2, 4, 9, 18, 19, 20 remains for cluster-2.

The results of FCM and KMean are almost similar. But as such the values of heterogeneity measures show that with application of 4 to 7 variables FCM has given consistency as well as acceptable homogeneity. HC method although shown the good demarcation of sites under different variables, but with increase in variables results did not remain consistent. SOM is a bit inferior to other methods but better than AP. As in this study instead of all variables prioritized variables are getting used stepwise, the attempt is made to get well demarcated clusters and good homogeneous sub-regions with as much as fewer variables utilized. So the FCM method with application of minimum 4 prioritised variables (1D,NR,SL,CA) are accepted as robust clustering method as well as optimum variables needed for clustering analysis. Very less change are seen with increase in variables as most of the variability is achieved with top 4 prioritised variables. . Recent studies have shown that a soft membership function is essential for finding high-quality clustering. Hardening the results obtained by fuzzy c-means algorithm produces better hard clustering solutions than those obtained by using the K-means algorithm (Hamerly and Elkan, 2002). Most catchments partly resemble to the properties of other catchments in the same cluster. The fuzzy cluster algorithm allows a catchment to have partial or distributed memberships in all the regions (Rao and Srinivas, 2006).

Again it is visualized that SH as a variable does not improve the clustering process in none of the methods. It is found that removing 2 or 3 stations from cluster-1 and cluster-2, homogeneous regions can be formed by all the methods but that will make clusters meaningless.

The clustering has been done successfully with at least minimum seven sites in a group. The final allocation of different sites to both the clusters with homogeneity measures is shown in Table 4.6. As finalized the FCM method with 4 variables result is accepted for final clusters. So cluster-1 with station ids 3, 5, 6, 7, 8, 10, 11, 14, 16 and cluster-2 with ids 1, 2, 4, 9, 12, 13, 15, 17, 18, 19, 20 are finally kept for further analysis. The calculation of growth period has been decided basing on 5T rule depending on the station year available in a particular cluster. Here cluster-1 has station year of 267and that of cluster-2 is 253. So dividing these values by 5, the present distribution can provide a better result up to 53 and 50 years return period for cluster 1 and 2. However a maximum of 50 year return period has been shown for comparison. The allotment of different sites under both cluster are shown in Fig. 4.6.

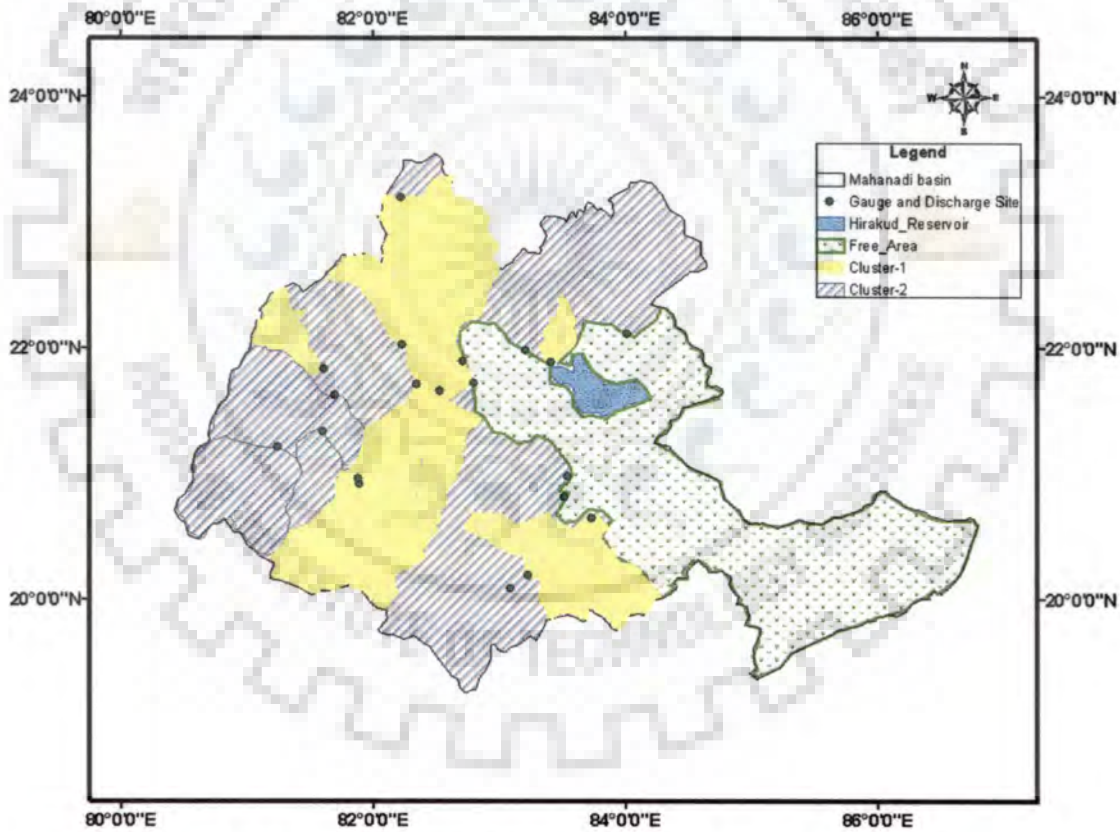


Fig. 4.6 Mahanadi basin divided to two clusters

For further analysis radar plot of all sites corresponding to 4 prioritised variables are given (Fig.4.7 (a) to 4.7 (d)). The radar plots show the distribution of variables to each site. It is revealed from radar plots that the sites 13, 1, 4, 9, 15 received maximum daily rainfalls during the years of study but the sites 12, 15, 18, 19, 20, 1, 2 are some of the sites with a bit high normal rainfall. The slopes of 14, 15, 18, 2, 7 are high to medium and the catchment area of 4, 5 and 14 are generally large than that of rest.

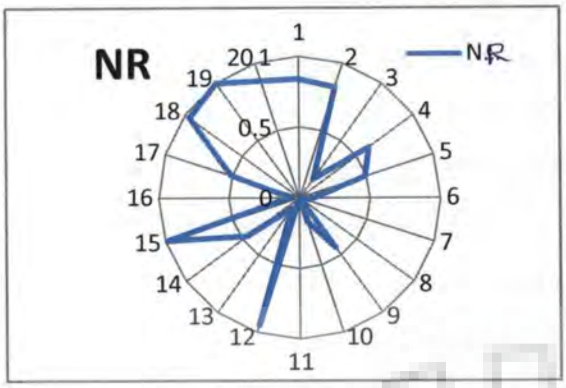


Fig. 4.7(a) Radar plot of Variable NR over 20 sites

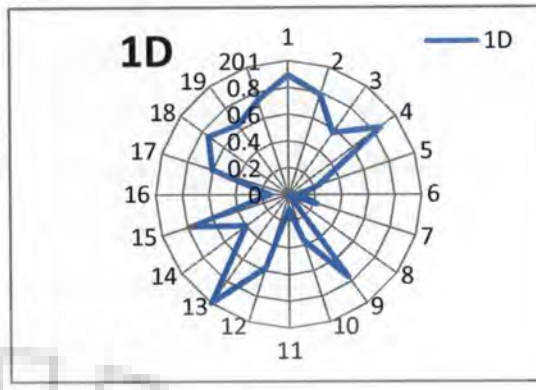


Fig. 4.7(b) Radar plot of Variable 1D over 20 sites

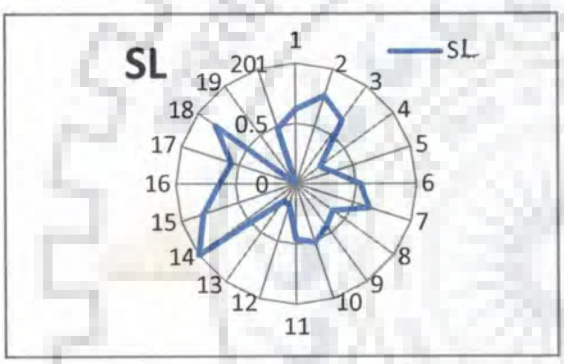


Fig. 4.7(c) Radar plot of Variable SL over 20 sites

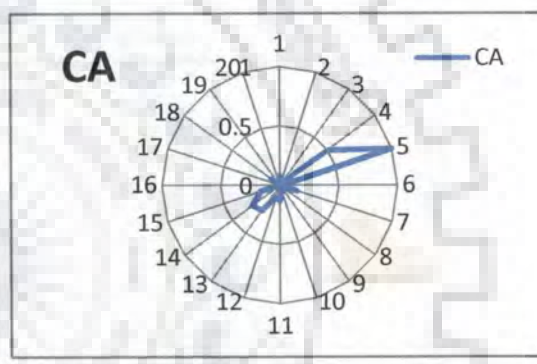


Fig. 4.7(d) Radar plot of Variable CA over 20 sites

From the box plot of both the clusters has been drawn in Figure 4.8(a) and 4.8(b). It has been concluded that the cluster-1 comprised of watersheds with low 1D, low NR, high to moderate SL, a low catchment area average with some big catchment as outlier. Cluster-2 represents high 1D, high NR, moderate to low slope and moderate catchment area average with outliers. The normalized average ID, NR, SL and CA for cluster-1 is 0.2355, 0.1683, 0.5713 and 0.1888. The corresponding average values of cluster-2 as 0.7605, 0.7337, 0.4644 and 0.1239. The DD and average maximum discharge (Q_m) are calculated separately for cluster-1 are found as 0.4048, 0.2713 and that of cluster-2 as 0.3219, 0.1848 respectively. So it becomes obvious that although cluster-1 has low rainfall characteristics, it has a high runoff generation capacity due to its slope and drainage density with respect to cluster-2.

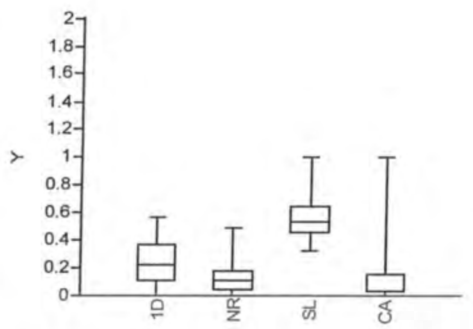


Fig. 4.8(a) Box-plot of variables of cluster-1

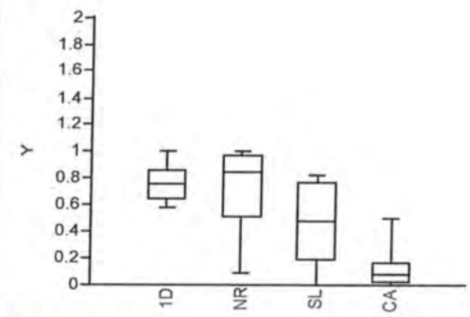


Fig. 4.8(b) Box-plot of variables of cluster-2

4.6.3 Development of Regional Flood Frequency Formulae for the Region

The development of regional flood frequency formulae requires following two relationships

(i) relationship between ($\frac{Q_t}{Q_m}$) and return period T

(ii) relationship between mean annual flood and catchment characteristics.

4.6.3.1 Relationship between ($\frac{Q_t}{Q_m}$) and return period T

The goodness of fit measure is considered on the basis of Z-statistics. The Z-statistics are shown in Table 4.6. which depicts that distributions like Pearson type-III(PT-III) and Generalised Pareto (GP) are suitable for cluster-1 and Generalised logistics (GL),Generalised extreme value (GEV), Generalised Normal (GN) and PT-III for cluster-2 as Z-values of these distributions are within -1.64 to + 1.64 .

Table 4.6 Final features of two clusters

Goodness of fit	Distribution	Cluster-1	Cluster-2
	GL	4.13	1.08
Z-statistics	GEV	2.59	**-0.30**
	GN	2.14	-0.53
	PT-III	1.25	-1.07
	GP	**-1.01**	-3.37

Considering the minimum Z-statistics it is observed that GP distribution is robust for cluster-1 and GEV for cluster-2. The regional parameters as observed from L-moment algorithm for the robust distribution are noted in Table 4.7.

Table 4.7 Regional parameters of different clusters

Regional parameters	Cluster-1	Cluster-2
Distribution	GP	GEV
ξ	0.157	0.744
α	1.089	0.429
κ	0.293	-0.021

4.6.3.2 Relationship between index flood and catchment characteristics

The regression relationships can be used for estimating index flood at ungauged sites (Brath et al. 2001). The index flood will act for both gauged and ungauged sites of the basin. For gauged sites mean or median values of annual peak series is taken as index flood and for ungauged sites in a hydrological homogeneous region, it is the multiple regression between mean or median annual flood and other independent catchment characteristics (eq.4.6 and 4.8). As there are 4 variables are selected as per PCA for clustering, the regression has been done keeping these variables only for finding a better index flood. It is found that the independent variable CA was deemed appropriate for either cluster as there was no significant development in R^2 was visible after adding more variables. Also for ready reckoner catchment area can be derived more easily. So the relationship between mean annual flood and catchment area is given (eq. 4.7 and 4.9) for two clusters using data of gauged catchments.

The mean discharge in m^3/s for both clusters in power form is as follows.

$$Q_{m1} = (589.33) X1D^{-0.197} XNR^{-0.8876} XSL^{0.2166} XCA^{0.9278} \tag{4.6}$$

$(R^2 = 0.863)$

$$Q_{m1} = (1.102) XCA^{0.877} \tag{4.7}$$

$(R^2 = 0.808)$

$$Q_{m2} = (3.8111 X 10^{-7}) X1D^{1.72} XNR^{0.929} XSL^{0.140} XCA^{0.7047} \tag{4.8}$$

$(R^2 = 0.785)$

$$Q_{m2} = (2.983) XCA^{0.761} \tag{4.9}$$

$(R^2 = 0.741)$

Where,

Q_{m1} = mean flood for cluster-1, Q_{m2} = mean flood for cluster-2, CA= catchment area and all the values are in their respective units.

The results so obtained are compared with the earlier studies made on Mahanadi basin. The Technical Report of National Institute of Hydrology, Roorkee, (1997-98) on sub-zone-3(d) regarding computation of growth factor on L-moment approach is compared with the growth factors obtained in this study. The comparative results are shown in Table-4.8. The subzone-3(d) covers 3 basins namely Mahanadi, Brahmani and Baitarani with a total catchment of around 193000 sq. km. The result of cluster-1 remains close to that of sub-zone-3(d) whereas cluster-2 growth factors remain less. So a justification has been made between sites to take a reasonable design flood without taking the result of sub-zone-3(d) as a gross.

Table 4.8 Growth factors for the two clusters compared to sub-zone-3(d)

Return period	Cluster-1	Cluster-2	NIH Report
	GP	GEV	Sub-Zone-3(d)
2	0.841	0.901	0.828
10	1.982	1.731	1.878
20	2.331	2.057	2.367
50	2.965	2.486	3.086

In another study (Singh and Seth, 1985) on Kelo bridge site Wakeby distribution has been applied for flood frequency analysis by using the available discharge series of (Gupta,1980) The same site (CA=1150 km^2) is also a site in this study (CA=1266 km^2) at Id No.16 although the position of both sites are different marginally. The Id No.16 falls into cluster-1 in this study. The growth factors obtained from observed values and that of Wakeby distribution has been compared with regional growth factor of cluster-1 in Table 4.9 and results shown are satisfactory. The observed values are quite close to result of GP distribution.

Table 4.9 Comparative growth factor for Kelo (Site Id No-16, Cluster-1)

Return period	This study	GUPTA	WAKEBY
	GP	(observed values)	distribution
2	0.841	0.851	0.877
10	1.982	2.026	2.100
20	2.331	2.457	2.612
50	2.965	2.992	3.293

Chowdhary (2005) has derived the equation for finding flood of return period 20 year (Q_{20}) and 50 year (Q_{50}) in his study on Mahanadi basin using index flood and multiple regression approach. The outcome of the study for 20 and 50 year return period index flood are as per eq.4.10 and 4.11.

$$Q_{20} = 1.94XQ_{IF}^{0.925} \tag{4.10}$$

$$Q_{50} = 4.074Q_{IF}^{0.829} \tag{4.11}$$

The index flood (Q_{IF}) is derived from the relationship with catchment area. The results are also compared with the result of this study (Fig. 4.9(a) and 4.9(b)). The results of the studies are plotted around 45^0 diagonal lines. It is found that the results of this study are more close to diagonal.

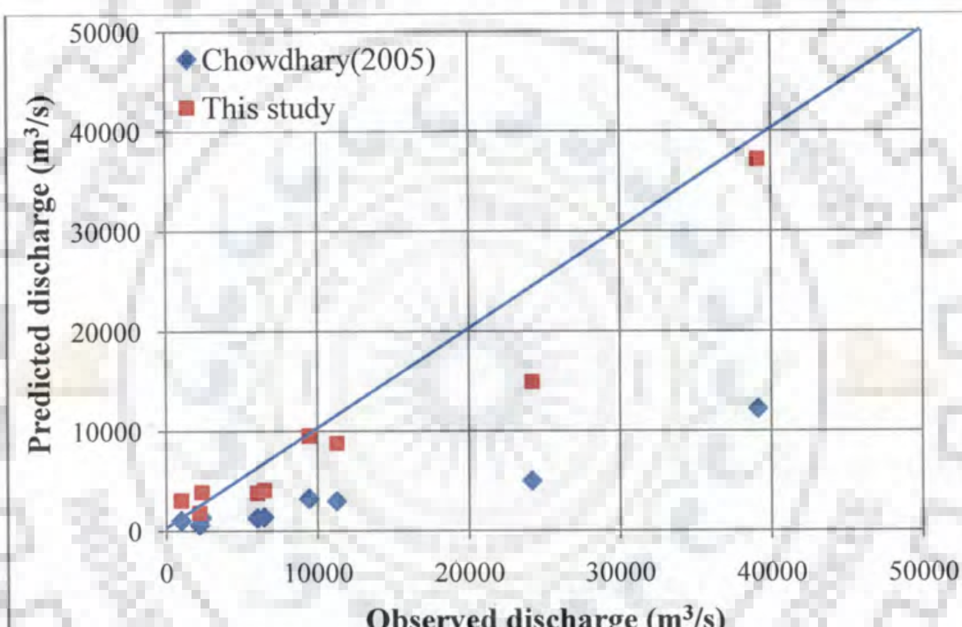


Fig. 4.9(a) Comparison of 50 year design flood in cluster-1

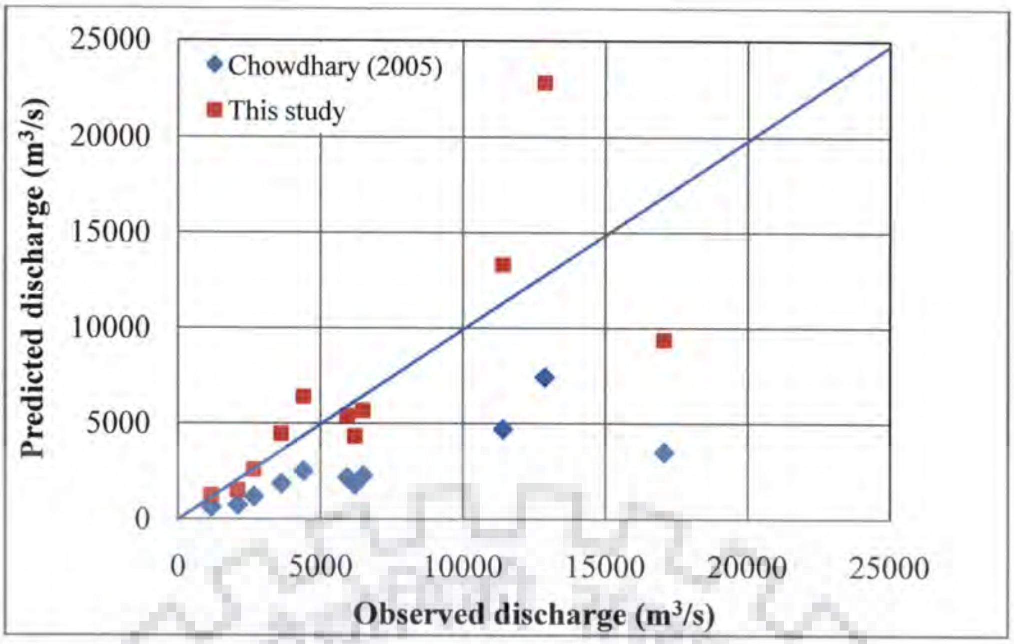


Fig. 4.9(b) Comparison of 50 year design flood in cluster-2

In comparison to previous studies on Mahanadi basin in this study the large basin is clustered into two homogeneous regions and the regional distributions and growth factors are fitted according to the sub-region. Earlier studies have taken Mahanadi or subzone-3(d) as one region to fit the regional equations or growth factors. The variable selections also represent the sub regions physically and hydro-meteorologically. In the process of clustering the physical and hydro-meteorological variables are considered. The results of clustering are also cross checked using number of methods and changing the variables.

4.7 CONCLUSIONS

The study describes the division of Mahanadi basin into two homogeneous clusters for making flood frequency analysis using optimum number of prioritized variables. The study also displays how the prioritized variables influence the clustering process. During prioritization of variables using PCA both statistical and physical importance of the variables towards runoff generation are considered. The following conclusions have been drawn from this study:

(i) FCM clustering method was found to be robust.

(ii) The SOM and AP are helpful in deciding the number of clusters.

(iii) Reducing the dimensionality of variables by using 4 variables out of 7 available have not put any significant impact on homogeneity and cluster formation.

(iv) The Generalised Pareto (GP) distribution holds good for cluster-1 and it contains the areas which can contribute substantially towards runoff generation due to high slope and drainage density characteristics.

(v) The cluster-2 contains areas with low runoff generation capacity as compared to cluster-1. Generalised Extreme Value (GEV) is the robust distribution for this cluster.

4.8 SCOPE FOR FURTHER WORK

The study has a wide scope for further work in the area. The possible directions in which further work can be undertaken are listed below:

i) The regional flood frequency analysis has been carried out based on the prioritized variables and clustering methods. In the present study only few variables have been considered. The other variables influencing runoff may be considered in future studies.

ii) The result of regional flood frequency analysis may be improved further by including data of more number of stations analysis more length of data.

iii) The regional flood frequency may be extended using partial duration series approach.

CHAPTER 5 - DEVELOPMENT OF FLOOD FORECASTING MODEL FOR DOWNSTREAM OF HIRAKUD RESERVOIR

5.1 BACKGROUND

Flood forecasting system should meet the requirements of accuracy, reliability, and timeliness. In recent years soft computing approaches such as artificial neural network, and fuzzy logic have been used in hydrological modeling and flood forecasting. In the present study soft computing techniques have been used for flood forecasting in Lower Mahanadi basin.

Presently, time lag method is being used by Department of Water Resources, Government of Orissa for forecasting the floods at different forecasting stations in the Mahanadi river system. In the present study the Mahanadi system downstream of Hirakud is modeled using traditional regression method, MLFF-ANN, RBF-ANN and TS-Fuzzy methods using the data of peak as well as 3-hourly discharge data of Khairmal, Barmul and Mundali sites. The results of applications have been inter-compared. The findings of this study may be applied in the field for strengthening the existing flood forecasting system.

5.2 INTRODUCTION

During the past few decades, a great deal of research has been devoted to the modeling and forecasting of river flow dynamics. Such efforts have led to the formulation of a wide variety of approaches and the development of large number of models. The existing models for river flow forecasting may broadly be grouped under two main categories namely, physically based models and black-box models. In between these two broad categories of models there is a wide variety of models e.g. deterministic and stochastic, lumped and distributed, event driven and continuous or their combinations.

The physical based models working on the principle of rainfall-runoff model require further information on evaporation, soil moisture, temperature and dam release etc. Sometimes all such information are not available on real time basis. It makes the application of physical based models for real time flood forecasting slightly difficult. In this context black box or data-driven models, which can discover relationships from input-output data without having the complete physical understanding of the system, may be preferable. While such models do not provide any information on the physics of the hydrologic processes, they are in particular, very useful for river flood forecasting where the main concern is accurate predictions of a flood level or discharge at specific river location.

In recent years, soft computing techniques, basically artificial neural networks and fuzzy logic approaches have been used in solving flood forecasting and other hydrological problems. A brief review of the applications of these techniques in flood forecasting is given in next section.

5.3 REVIEW OF LITERATURE ON APPLICATION OF SOFT COMPUTING TCHNIQUES IN FLOOD FORECASTING

5.3.1 Artificial Neural Network (ANN)

ANNs are basically data driven approaches and are considered as black box models (Bishop, 1994) in hydrological context. These models are capable of adopting the non-linear relationship (Flood and Kartam, 1994) between rainfall and runoff as compared to conventional techniques, which assume a linear relationship between rainfall and runoff. ANNs have strong generalization ability, which means that once they have been properly trained, they are able to provide accurate results even for cases they have never experienced before (Imrie et al., 2000).

Over the last decades, Artificial Neural Networks (ANNs) have been increasingly used in hydrological forecasting (Maier and Dandy, 2000). Previous studies have shown that ANNs are capable of reproducing unknown rainfall-runoff relationship adequately (ASCE 2000a, ASCE, 2000b). ANN is also a powerful tool in solving complex nonlinear river flow forecasting problems (Hsu et al., 1995, 2002; Thirumalaiah and Deo, 1998; Atiya et al., 1999; Uvo et al., 2000; Birkundavyi et al., 2002) and in particular when the time required generating a forecast is very short.

Jayawardena et al. (1996) compared the performance of both MLP and RBF networks for an experimental drainage basin at China to predict water levels. The RBF models resulted in better performance and took less time for model development than MLP as no repetition is required to reach the optimum model parameter. Minns and Hall (1996) applied for rainfall-runoff modeling. They have applied ANN for both one and two hidden layers and found that the results improved with two hidden layer but the extra computational effort does not seem justified.

Dawson and Wilby (1998) discussed the application of ANNs to flow forecasting in two flood-prone catchments in England using hourly hydrometric data.

Zealand et al. (1999) explores the capabilities of ANNs and compares the performance of this tool to conventional approaches used to forecast streamflow. Several issues associated with the use of an ANN are examined including the type of input data and the number, and the size of hidden layer(s) to be included in the network. Campolo et al. (1999) developed the NN model to forecast the water level using the distributed rainfall.

Liong et al. (2000) achieved to get a high degree of accuracy with ANNs for river stage forecasting in Bangladesh. Coulibaly et al. (2000) applied MLFF network for reservoir inflow forecasting with Levenberg–Marquardt back propagation (LMBP) and cross-validation technique to avoid underfitting or overfitting on FNN training and enhances generalization performance. Imrie et al. (2000) applied cascade correlation learning architecture for river flow prediction.

Chang et al. (2001) developed a rainfall-runoff model for 3-hour ahead flood forecasting using a two stage unsupervised and supervised learning. They recommended that the RBFNN can be considered as a suitable technique for flood flow.
Rajurkar et al. (2002) used a linear multiple-input-single-output (MISO) model coupled with ANN for developing rainfall-runoff relationship for river Narmada, India. The model provided a systematic approach for runoff estimation.

Ni and Xue (2003) established an ANN model based on Radial-Basis-Function (RBF) for flood risk ranking at five safety polders in Yangtze River, China. The authors developed site-specific and multi-site-specific RBFNN model to provide useful tools for rapid prediction of flood routing process. It also shows much promise for rapid feedback of the site-specific risk and real-time diversion process control. Suitability of some deterministic and statistical techniques along with an ANN to model an event based rainfall-runoff process have been investigated by Jain and Indurthy (2003).

Lekkas et al. (2004) utilizes various types of ANNs in an attempt to assess the relative performance of existing models. Ali Efenti, a sub-catchment of the river Pinios (Greece), is examined and the results support the hypothesis that ANNs can produce qualitative forecasts. A 7-hour ahead forecast in particular proves to be of fairly high precision, especially when an error prediction technique is introduced to the ANN models. Solomatine and Xue (2004) presented an approach to building modular rainfall-runoff models where, based on expert judgment encapsulated in simple rules, input data was partitioned into several subsets, and separate ANN or M5 tree models were built for each subset. Kisi (2004) demonstrated the application of ANNs in predicting mean monthly streamflow. The estimation of flow is very important for reservoir operation policy; therefore, in this study, based on the monthly flow data obtained from the Turkey State of Water Works, ANNs have been used to predict river flow. Autoregressive (AR) models have also been applied to the same data. The performance of the neural network is compared with that of the statistical method.

Senthil Kumar et al. (2005) compared rainfall-runoff modeling skills of two ANN configurations MLP and RBF networks. The results of their study indicate that the generalization properties of RBF networks are poor compared with those of MLPs in rainfall-runoff modelling. Bhattacharya and Solomatine (2005) built water level-discharge relationship models with an ANN and a M5 model tree on the river Bhagirathi in India. Based on casual meteorological parameters, Ahmad and Simonovic (2005) applied ANNs for predicting the peak flow, timing and shape of runoff hydrograph for the Red River in Manitoba, Canada.

Chau (2006) adopted a particle swarm optimization model to train an ANN model to predict water levels in Shing Mun River of Hong Kong with different lead times on the basis of upstream gauging stations. Tareghian and Kashefipour (2006) developed an ANN model to forecast daily runoff for Karoon River in Iran. Dawson et al. (2006) used ANNs to predict T-year flood events and the index flood for 850 catchments across the UK. Wu and Chau (2006) employed a genetic algorithm based Artificial Neural Network (ANN-GA) for flood forecasting in a channel reach of the Yangtze River in China. Sahoo and Ray (2006) demonstrated that the ANN can outperform rating curves for discharge forecasting.

Kisi (2007) presented a comparison of different artificial neural networks (ANNs) algorithms for short term daily streamflow forecasting. Four different ANN algorithms, namely, backpropagation, conjugate gradient, cascade correlation, and Levenberg–Marquardt are applied to continuous streamflow data of the North Platte River in the United States. The models are verified with untrained data. The results from the different algorithms are compared with each other. The correlation analysis was used in the study and found to be useful for determining appropriate input vectors to the ANNs. Pang et al. (2007) developed a Nonlinear Perturbation Model (NLPM) based on ANN, defined as NLPM-ANN, for the purpose of improving the rainfall-runoff forecasting efficiency and accuracy. Chang et al. (2007) presented a systematic investigation of the three common types of ANNs for Multi-Step-Ahead (MSA) flood forecasting for two watersheds in Taiwan. Kisi and Cigizoglu (2007) explored the use of MLP ANNs, radial basis ANNs, and GRNNs for forecasting intermittent flow series.

Turan and Yurdusev (2008) applied FFBPNN and GRNN along with Mamdani Fuzzy logic and regression models on four runoff gauging stations on Birs river basin of Switzerland for flow prediction at downstream location.

Mukerji et al. (2009) compared relative performance of ANN, ANFIS and adaptive neuro-GA integrated system (ANGIS) models in flood forecasting. They used various performance measures to evaluate and compare the performances of different models. They concluded that the performance of ANGIS model is better than ANN model in most cases.

Wu et al. (2010) proposed modular artificial neural network (MANN) model for daily and monthly rainfall forecasting. MANN model was developed by portioning the training data into 3 clusters by the FCM technique and then each subset was approximated by a single NN model. The final output of the model resulted directly from the output of the three local models. MANN was compared with three benchmark models viz. NN, K-nearest-neighbours (K-NN) and linear regression (LR). Results of models without preprocessed inputs indicated that MANN model performed the best among all four models.

Tiwari and Chatterjee (2010) developed a hybrid wavelet-bootstrap-ANN (WBANN) model to explore the potential of wavelet and bootstrapping techniques to develop an accurate and reliable ANN model for hourly flood forecasting. The wavelet technique decomposes the time series data into different components capturing useful information at various resolutions. The model is applied in Mahanadi basin and the results are compared with 3 different ANN models viz. traditional ANNs, wavelet based ANNs (WANNs) and bootstrap based ANNs (BANNs). The overall performance of WBANN model is accurate and reliable than other 3 models.

Deshmukh and Ghatol (2010) applied static neural approach by applying RBF to rainfall-runoff modeling for the upper area of Wardha river in India. They recommended the Levenberg-Marquardt learning rule and Tanh activation function is a more versatile combination for RBF network.

5.3.2 Fuzzy Logic

Fuzzy logic, another soft computing method, has emerged as a convincing alternative to traditional procedures in the analysis and prediction of various real world phenomena. It has also opened up new avenues to hydrological modeling and it comes only after Zadeh (1965) explored it first.

Since Zadeh (1965) publication regarding an extension of the classical fuzzy set theory, the fuzzy method has been widely used in many fields of applications, such as pattern recognition, data analysis, system control, etc. (Kruse et al., 1994; Theodoridis and Koutroumbas, 1999).

See and Openshaw (1999) indicated that the fuzzy logic can be used in combination with other soft computing techniques to create sophisticated river level monitoring and forecasting system.

Hundecha et al. (2001) demonstrated that a Fuzzy Logic (FL) approach could be used to simulate actual component hydrologic processes in areas where sufficient data were not available to model these processes physically. Ozelkan and Duckstein (2001) proposed a fuzzy conceptual rainfall-runoff framework to deal with parameter uncertainties of conceptual rainfall-runoff models.

Cheng et al. (2002) combined a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff Xinanjiang model calibration in the Shuangpai reservoir.

Luchetta and Manetti (2003) used a fuzzy clustering approach to forecast a real time hydrological model in the Padule di Fucecchio basin in middle-north of Italy. Mahabir et al. (2003) applied FL to forecast seasonal runoff for the Lodge and Middle Creek basins, Canada.

Blazkova and Beven (2004) used FL to estimate flood frequency by continuous simulation of sub-catchments rainfalls and discharges for a dam site in a large catchment in the Czech Republic.

Vernieuwe et al. (2005) described the catchment's response to rainfall input through fuzzy relationships for Zwalm River in Belgium.

Rao and Srinivas (2006) tested a fuzzy clustering for regionalization of watersheds with 245 gauging stations data in Indiana. Jacquin and Shamseldin (2006) developed TS fuzzy model and applied to six catchment of diverse climatic characteristics and reported that fuzzy model is better than simple liner model, linear perturbation model, and nearest neighbor perturbation model. They reported that fuzzy inference systems are suitable alternative to the traditional methods for modeling nonlinear relationship between rainfall and runoff processes.

Nayak (2010) developed a popular fuzzy rule-based model for river flow forecasting for Anandapur gauging station of the Baitarani basin in India. The developed model is used to forecast up to 12 hour in advance. It is observed that the developed model follows the trend of the input membership grade in antecedent part of the fuzzy model.

Lohani et al. (2011) developed daily rainfall runoff models using artificial neural network (ANN), fuzzy logic (FL) and linear transfer function (LTF)-based approaches for upper Narmada basin, India. They compared these models with different input data vectors. This study further investigates the impact of antecedent soil moisture conditions in the performance of the daily rainfall-runoff models. The results show that the fuzzy modelling approach is uniformly outperforming the LTF and also always superior to the ANN-based models. Gogoi and Joshi (2011) applied a simple form of fuzzy logic based modeling, based on the theory of fuzzy sets is implemented to forecast flood in Jiadhal river basin of Assam.

In Mahanadi basin the discharge data are available at 3 hour interval during the monsoon period. However, sometimes due to various reasons only peak discharge data are available at base and / or forecasting stations. Therefore, in order to provide reliable flood forecasts to the flood managers and district administration the following two types of input data vectors have been used:

(i) peak discharge, and

(ii) 3-hourly discharge

This chapter deals with development of flood forecasting models using both continuous (3- hourly) and discrete (peak floods) discharges at different G&D stations at downstream of Hirakud. The following models have been attempted.

i) Statistical model,

ii) ANN model based on MLFF and RBF network

iii) Fuzzy logic model based on Takagi–Sugeno principle

5.4 STUDY AREA

In this study flood forecasting is confined to the downstream of Hirakud reservoir. The base station Khairmal is about 115 km. downstream of Hirakud. The schematic diagram of downstream Mahanadi basin is shown in Fig. 2.3 (Chapter 2). The release from Hirakud reservoir along with the contribution of intercepted catchment between dam and Khairmal remains the prime input for the model. The intermediate station is at Barmul. The forecasting station Mundali is just before the starting of delta region. The location of these stations is shown in Fig.5.1.

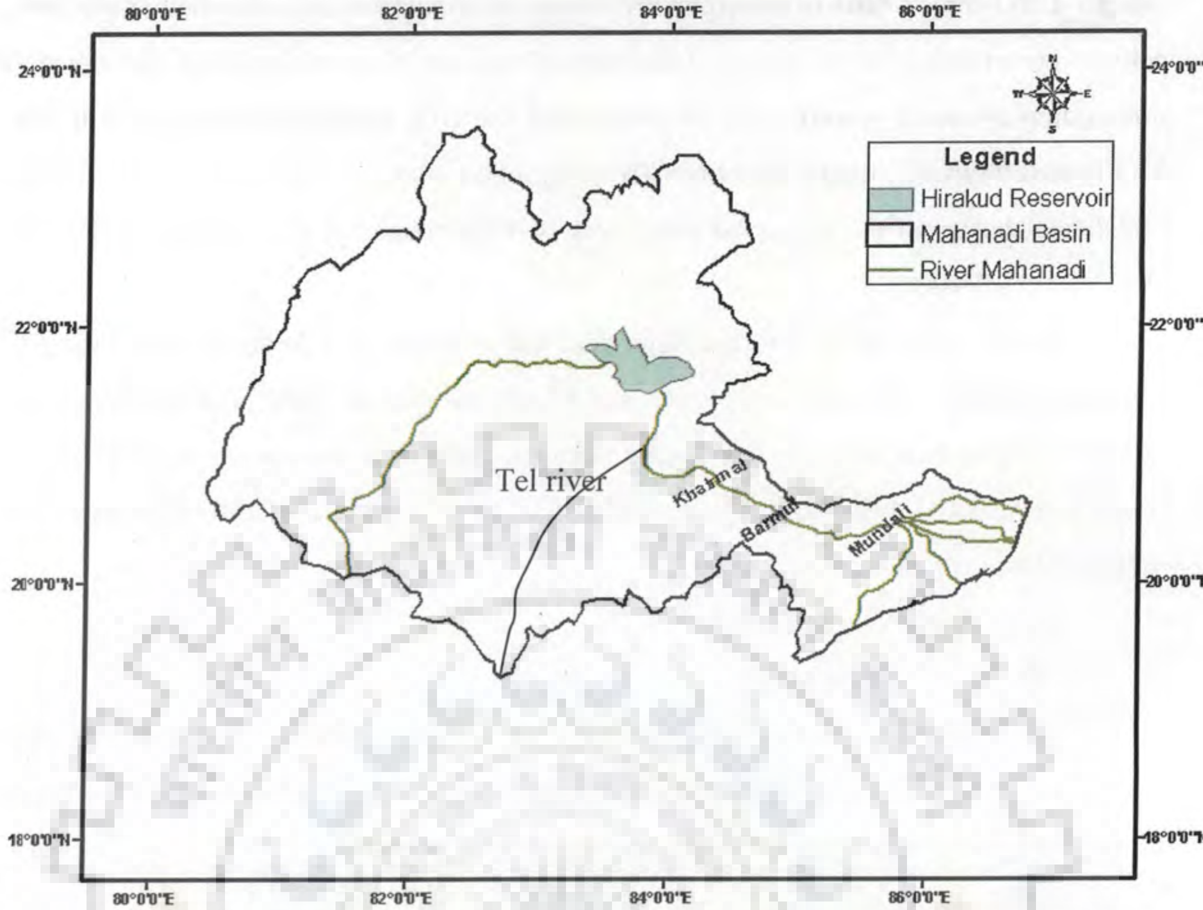


Fig. 5.1 Mahanadi basin map showing the base, intermediate and forecasting station

5.5 DATA AVAILABILITY

The discharge data at 3-hourly interval is available from 1996-2010 for the stations at Khairmal, Barmul and Mundali. For peak discharge a total of 101 peak values are selected and out of this 51 peaks are selected for calibration and remaining 50 are used for validation. The peak values and then corresponding travel times are derived from the past historical data.

Total 13080 values of 3-hourly discharges have been collected. The discharge from 1996-2003 (6666 values) are used for calibration and data from 2004-2010 (6414 values) are used for validation.

5.6 METHODOLOGY

5.6.1 Statistical Model

In statistical model correlation (either simple regression or multiple regressions) between stage or discharges of upstream and downstream gauging stations are established. It is one of the simplest methods to forecast the flood at downstream sites. This method gives better results when the contribution of tributaries joining the main stream in between the base station an forecast station is not significant and the catchment is not receiving heavy rainfall.

5.6.1.1 Linear Transfer Function (LTF)

The general representation of statistical model (multiple linear regression) is given by

$$Y_i = \sum_{j=0}^{k} \beta_j X_{ij} + \varepsilon \tag{5.1}$$

With $X_{i0} = 1$. Here, X_{ij} is the independent variable for the i^{th} observation (various discharge values in the present study), Y_i is the dependent variable for the i^{th} observation, β_j is unknown coefficients to be estimate, k is the number of coefficient (to be estimated) in the model, and ε is the error in the determination of Y_i which is generally assumed as having zero mean and constant standard deviation σ. The unknown coefficients (β) are estimated by least square method.

5.6.2 ANN Model

Artificial neural networks (ANN) are simplified models of the biological neuron system and as such they share some advantages that biological organisms have over standard computational systems. A typical ANN model consists of number of layers and nodes that are organized to a particular structure (Mehrotra et al., 1997).

There are various ways to classify a neural network. Neurons are usually arranged in several layers and this arrangement is referred to as the architecture of a neural net. Networks with several layers are called multi layer networks as opposed to single layer networks that only have one layer. The classification of neural networks is done by the number of layers, the direction of information flow, the non-linear equation used to get the outputs from the nodes and the method of determining the weights between nodes of different layers. Within and among the layers the neurons can be connected in two basic ways:

(1) *Feedforward networks* in which the neurons are arranged in several layers. Information flows only in one direction, from the input layer to output layer, and
(2) *Recurrent networks* in which neurons are arranged in one or more layers and feedback is introduced either internally in the neurons to other neurons in the same layer or to neurons in preceding layers.

The commonly used neural network is three-layered feed forward network due to its general applicability to a variety of different problems (Hsu et al.1995)

5.6.2.1 Multiple Layered Feed Forward Network (MLFF or MLP)

Feed forward ANN comprises of a system of units analogous to neurons, which are arranged in layers. Between the input and output layers there may be one or more hidden layers. The units in each layer are connected to the units in the subsequent layer by a connection weight w , which is adjusted during training. A data pattern comprising of the values x_i present in the input layer i is propagated forward through the network towards the first hidden layer j . Each hidden unit receives the weighted outputs from the units of the previous layer. These are summed to produce a net value, which is then transformed to an output value using an activation function. Fig. 5.2 represents the components of ANN architecture along with its processing unit. ANN models should be trained properly before using it for testing. Generally, training data patterns are fed sequentially into the input layer and this information is propagated through the network. The resulting outputs $y_j(t)$ are compared with the corresponding desired or actual output, $d_j(t)$. The mean squared error at anytime $t(E(t))$ is calculated over the entire data set using Eq. (5.2). The intermediate weights are adjusted using an appropriate learning rule until $E(t)$ has decayed to a suitable level.

$$E(t)=\frac{1}{2}*\sum(y_j(t)-d_j(t))^2 \tag{5.2}$$

Numerous algorithms have been developed for training of ANN models to attain optimum model performance by ensuring good generalization and computational efficiency. For feed forward ANN, the back propagation algorithm is frequently used for ANN training. Training of ANN using error back propagation algorithm proceeds as follows. The weights, w_{ji} are adjusted during training using the following equation:

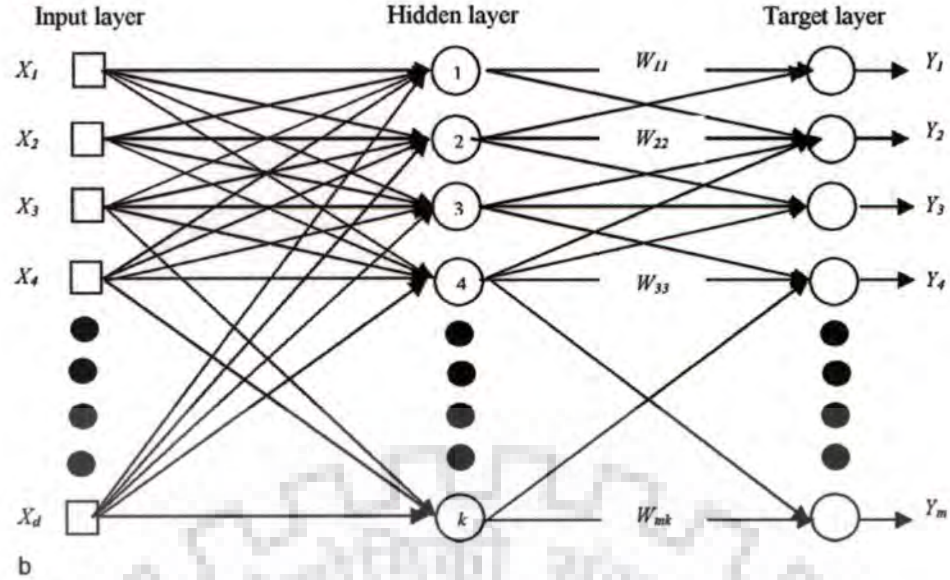
$$w_{ij}(t+1)=w_{ji}(t)+\Delta w_{ji}(t+1)+\mu\Delta w_{ji}(t) \tag{5.3}$$

where, μ is the momentum constant. The weight increment Δw_{ji} may be found using the gradient descent method.

$$\Delta w_{ji}=\eta\frac{\partial E}{\partial w} \tag{5.4}$$

where, η is the learning rate. The gradient descent method makes the changes to the weights in the direction of the steepest descent down the error surface by reducing the likelihood that the network will stabilize at a local rather than a global error minimum. The momentum term in Eq. (5.4) adds inertia to the training procedure. With an increase in the learning rate, the error gradient gets smaller and thereby helps to avoid the entrapment in local minima.

a



b

Fig. 5.2 ANN model structure

5.6.2.2 Radial Basis Function (RBF)

Although the MLP can produce accurate forecasts, it does have a number of drawbacks. For instance, training an appropriate MLP can take a long time and it has number of parameters that to be determined by the user. On the other hand, the RBF has been used in a limited number of studies (Mason et al., 1996; Fernando and Jayawardena, 1998; Dawson and Wilby, 2001), it can be trained in a fraction of the time, has fewer parameters to be determined, and, in certain cases, predicts river flow more accurately than the MLP (Fernando and Jayawardena, 1998; Sudheer et al., 2002b)

An alternative to the MLP is the Radial Basis Function (RBF) network (Bianchini et al., 1995; Chen et al, 1991) which has linear parameters and has found applications in other areas such as electrical and electronic engineering. Park & Sandberg (1991) proved theoretically that the RBF type ANNs are capable of universal approximations and learning without local minima, thereby guaranteeing convergence to globally optimum parameters. For hypothetical situations, Moody & Darken (1989) demonstrated that the RBF type networks learn faster than MLP networks. This study attempts to apply the RBF approach to real situations of flood water level predictions and to compare the model performances with those of MLP network models.

The Radial Basis Function Neural networks have a very strong mathematical foundation rooted in regularization theory for solving ill-conditioned problems. An RBF network is a three-layer feed-forward type network in which the input is transformed by the basic functions at the hidden layer. At the output layer, linear combinations of the hidden layer node responses are added to form the output. The mapping function radial basis function network is built up of Gaussians rather than sigmoids as in MLP networks. Learning in RBF network is carried out in two phases: first for the hidden layer, and then for the output layer. The hidden layer is self-organising and its parameters depend on the distribution of the inputs, not on the mapping from the input to the output. The name RBF comes from the fact that the basic functions in the hidden layer nodes are radially symmetric. The output layer, on the other hand, uses supervised learning (gradient or linear regression) to set its parameters.

Chen et al. (1991) reported that the choice of the basis function is not crucial to the performance of the network. The most common choice however, is the Gaussian function which can be defined by a mean and a standard deviation. Figure 6.3 shows a schematic diagram of an RBF network with input (N), hidden (L) and output layer (M) nodes for the general transformation of ND points of $X(X......X......X^{ND})$ in the input space to points $Y(t......F......F^{ND})$ in the output space. The parameters of an RBF type neural network are the centers and the spreads of the basis functions at the hidden layer nodes and the synaptic weights w_{kj} of the output layer nodes. The RBF centers are also points in the input space. The basis function response for an input depends on the distance between the point representing the input (X) and the RBF centre (U_j).

For an input X^i, the j^{th} hidden node produces a response h_j given by,

$$h_j = \exp\left\{\frac{\left\|X^i - U_j\right\|}{2\sigma_j^2}\right\} \tag{5.5}$$

Where, $\left\|X^i - U_j\right\|$ represents the distance between input X^i and centre of j^{th} hidden node measured by Euclidean distance method.

The output is represented by $Y_{ik} = \sum_{j=1}^{L} h_j w_{kj}$ (5.6)

5.6.3 Fuzzy Inference System (FIS)

Fuzzy inference is the process of mapping a given set of inputs to an output using fuzzy logic defined by the fuzzy 'If-Then' rule database. The mapping provides a basis for computing the decisions. There are two types of FIS that can be implemented.

- Mamdani type
- Sugeno type

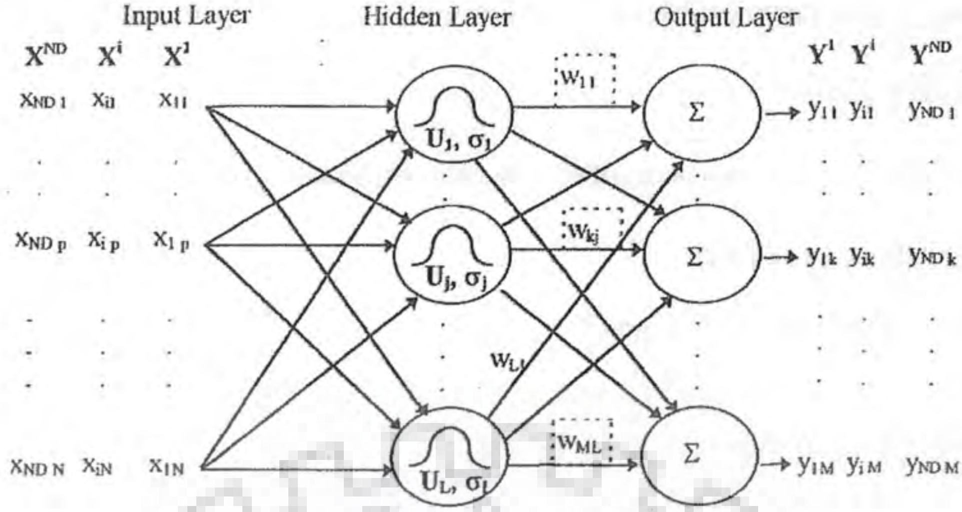


Fig. 5.3 Schematic diagram of RBF Network

In the rule bases described hitherto with the IF-THEN rules fuzzy sets both in the premises and in the conclusions are used. This kind of inference is called Mamdani inference. A modified inference scheme, developed by Takagi and Sugeno, represents the conclusions by functions. Recently Takagi Sugeno fuzzy approach is gaining popularity for solving non-linear problems. A rule of this form will be IF x_1 is A_{r1} AND x_2 is A_{r2} THEN $u = f_r(x_1, x_2 \ldots\ldots x_n)$ (5.7)

The structure of the premises is the same as for the Mamdani inference. However, in the conclusion all linguistic terms B_r are substituted by the functions f_r, and therefore it is not necessary to define a priori linguistic terms $B_r(u)$ for the conclusions.

The function f_r represents a direct mapping from the input space $(x_1 * x_2 * \ldots\ldots x_n)$ with the input values $x_1, x_2 \ldots\ldots x_n$ to the output space u.

$$u = f_r(x_1, x_2 \ldots\ldots x_n) \tag{5.8}$$

The connective operation in a rule is in this case performed via the degree of relevance H_r of the premise of the rule R_r and the function f_r in the conclusion. The final output is determined as a weighted mean value over all R rules according to

$$u' = \frac{\sum_R H_R f_r(x_1, x_2 \ldots\ldots x_n)}{\sum_R H_r} \tag{5.9}$$

The effort of performing a defuzzification is saved, as the crisp value u' is directly determined by the inference operation and this makes this method attractive.

The Takagi Sugeno fuzzy system builds an overall combination of functions f_r, which are valid in some range. If the membership functions of the fuzzy sets in the premises are overlapping, the transition between the functions is always continuous. For the special case of linear functions $f_r(x_1, x_2........x_n) = \sum_{v=1}^{n} c_{rv} x_v$ (5.10)

the coefficients c_{rv} can be determined by some identification procedure.

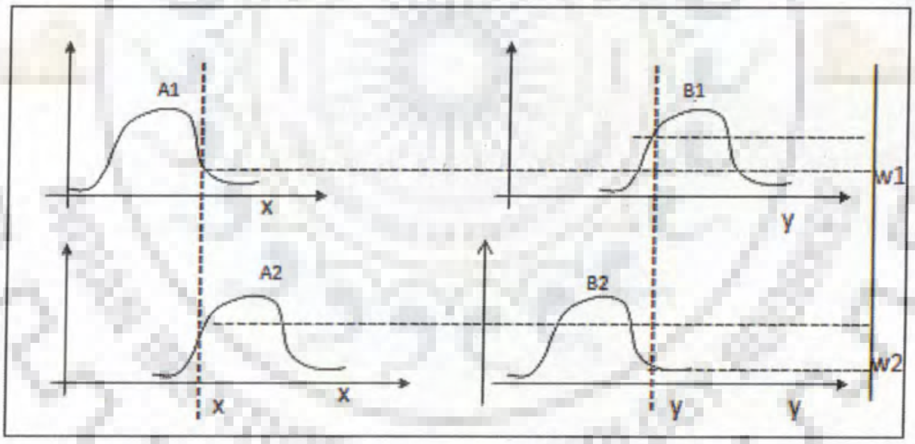
5.6.3.1 Architecture of FIS

Consider that the FIS has two inputs x, y and one output z. Fig. 5.4 and 5.5 illustrate a TS fuzzy inference system. For a first-order Takagi Sugeno (TS) model, a common rule set with two fuzzy if-then rules can be written as follows:

Rule 1, if x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$, and (5.11)

Rule 2, if x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$ (5.12)

Where the "if" statement is the antecedent, the "then" statement is the consequent, x and y are linguistic variables and A_1, A_2, B_1, B_2 are corresponding fuzzy sets, and p_1, q_1, r_1 and p_2, q_2, r_2 are linear parameters.



$f_1 = p_1x + q_1y + r_1$ $f_2 = p_2x + q_2y + r_2$

Fig. 5.4 FIS membership functions (MFs) and rule generation

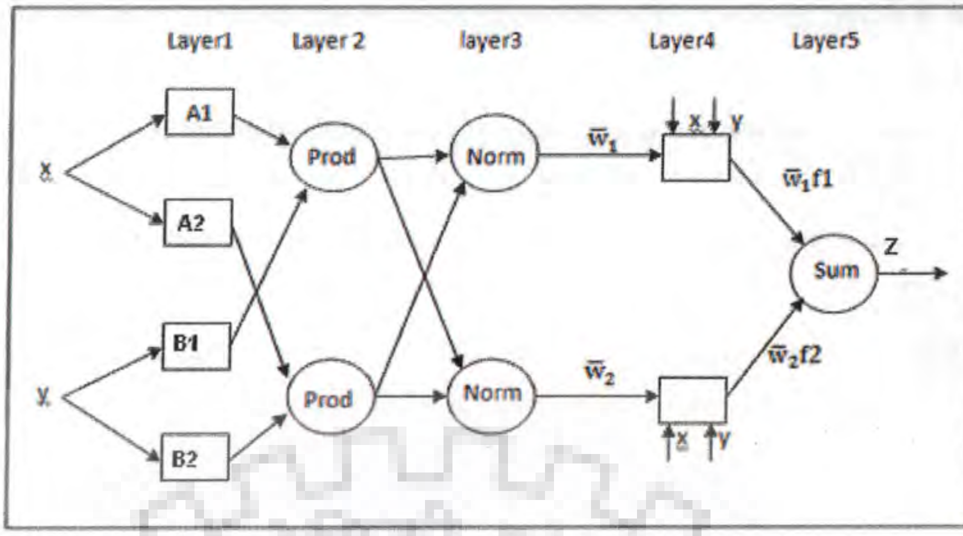


Fig. 5.5 FIS network

5.6.3.2 Fuzzy structure identification

Data driven fuzzy identification is an effective tool for the approximation of uncertain non-linear systems (Hellendoorn and Driankov, 1997). The core of the fuzzy structure identification method is in the clustering and the projection. First, the output space is partitioned using a fuzzy clustering algorithm. Second, the partitions (clusters) are projected onto the space of the input variables. The output partition and its corresponding input partitions are the consequents and antecedents, respectively. Then by projecting each cluster onto each input variable, temporary clusters in the input space are obtained. This may be implemented by using the subtractive clustering method that automatically determines the number of cluster. The subtractive clustering method uses the following formula to express the potential as a sum of contribution of Euclidean distance between a given point and all other data points (Chiu, 1994):

$$D_i = \sum_{j=1}^{N} d_{ij} \tag{5.13}$$

$$d_{ij} = e^{-\alpha |x_i - x_j|^2}, \qquad i = 1, 2........N \tag{5.14}$$

where D_i is the potential of the data point x_i to be a cluster centre, d_{ij} denotes the contribution of every single distance, N is the number of training data samples and $\alpha = 4/r_a^2$ when r_a is the cluster radii.

The clustering algorithms are used extensively not only to organize and categorize data, but are also useful for data compression and model construction. Clustering partitions a data set into several groups such that the similarity within a group is larger than that among groups. The clustering techniques are validated on the basis of the two assumptions viz.

(i) similar inputs to the target system to be modelled should produce similar ouputs, and

(ii) These similar input-output pairs are bundled into clusters in the training data set.

The subtractive clustering method assumes that each data point is a potential cluster center and calculates a measure of the likelihood that each data point would define the cluster center, based on the density of surrounding data points. The subtractive clustering algorithm performs the following tasks

(i) selects the data point with the highest potential to be the first cluster centre,

(ii) removes all data points in the vicinity of the first cluster centre (as determined by radii), in order to determine the next data cluster and its centre location and

(iii) iterates on this process until all of the data is within radii of cluster centre.

5.6.4 Model Development

The flood forecasting models have been developed on peak as well as 3-hourly discharge values of base station (B.S) and intermediate station (I.S). Therefore, initially the peak values of B.S. and its corresponding peaks at downstream intermediate and forecasting stations (F.S) are derived from the time series data of river discharge. These peak values are used as inputs and target outputs for the ANN, Fuzzy as well as statistical model. The travel times of various flood peaks have been computed and grouped in different clusters by using fuzzy c-means, k-means techniques. The average travel times are fixed for average discharge of each cluster.

For continuous discharges a possible number of alternative inputs are proposed to be tested in all networks of ANN and fuzzy. The input vectors are selected generally by trial and error method (Maier and Dandy, 2000). Determinations of number of discharge values involve computation of lags of discharge values that have significance influence on forecasted flow. These influencing values corresponding to different lags can be identified through statistical analysis of the data series by avoiding the trial and error procedure.

Determining the number of flow values involves the computation of lags of flow values of different time interval that have significance influence on forecasted discharge. In this study, the number of parameters corresponding to different antecedents was determined by two statistical methods, i.e. autocorrelation function (ACF) and partial autocorrelation function (PACF) between the variables. The ACF and PACF are generally used to gather information about the autoregressive process of the data series (Sudheer et al., 2002). The number of antecedent river flow values that should be included in the input variables is usually determined by placing a 95% confidence interval on the autocorrelation and partial autocorrelation plots. The base station flows have significant relation with the flow of forecasting station. The statistical parameters such as auto correlation function (ACF), partial auto correlation function (PACF) and cross correlation function (CCF) can be used for this purpose.

The cross-correlation is a measure of similarity of two signals. It is a function of the relative time between the signals. The cross correlation coefficient between rainfall and runoff, which was calculated by normalizing the cross-correlation of the two signals, has been used to identify the time lag (offset) where the similarity is highest (Teschl and Randeu, 2006).

Therefore on the basis of PACF and CCF of the data series, the input vectors have been selected for the flood forecasting model in the present study.

5.6.5 Performance Criteria

One major problem in assessing ANN models is the use of global statistics (RMSE, Coefficient of correlation, Efficiency, etc.) in calibration. When this approach is employed for modeling studies, the solution will, in most cases, produce a higher or near perfect 'goodness of fit' statistic. The statistical indices considered are the root mean square error (RMSE), coefficient of correlation (R) and the model efficiency, which are defined as follows:

5.6.5.1Coefficient of correlation (R)/ Coefficient of determination (R^2)

Coefficient of correlation (R) is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information. It provides a measure of how well future outcomes are likely to be predicted by the model.

$$\text{Coefficient of correlation (R)} = \frac{\sum_{i=1}^{N}(Q_o-\bar{Q}_o)(Q_p-\bar{Q}_p)}{\sqrt{\sum_{i=1}^{N}(Q_o-\bar{Q}_o)^2\sum_{i=1}^{N}(Q_p-\bar{Q}_p)^2}} \qquad (5.15)$$

Coefficient of correlation is a statistic that will give some information about the goodness of fit of a model. In regression, the Coefficient of correlation is a statistical measure of how well the regression line approximates the real data points. R value of 1.0 indicates that the regression line perfectly fits the data.

5.6.6 Model parsimony

As far as model inputs are concerned the parsimony of the models should be observed by using optimum number of inputs. The performance of calibration and validation is highly dependent on the structure of the model. A model having large structure have difficulty in converging during training or may over fit whereas a model having less parameters may not represent the underlying physical processes perfectly.

Anders and Korn (1999) proposed the use of information criteria to compare different ANN models. The idea behind the use of information criteria is to find an optimal trade-off between an unbiased approximation of the underlying model and the loss of accuracy caused by estimating a number of parameters. A variety of different criteria can be found in the literature (Sudheer, 2000). The most prominent, and still widely used, criterion is the Akaike Information Criterion (AIC), proposed by Akaike (1974). Shibata (1976) showed that a consistent estimator is not found using the Akaike Information Criteria (AIC) and modifications to this criterion have been proposed, such as the Bayesian Information Criteria (BIC).

According to Shumway (1988), the BIC leads most often to a correct model structure and has the smallest prediction error. Senthil Kumar et al. (2005) have successfully applied AIC and BIC in their rainfall-runoff model to select the best model. In the current analysis, both these criteria have been used for comparing the performance of ANN and fuzzy models. According to Hsu et al. (1995) the AIC and BIC can be computed using the following equations:

$$AIC = m\ln(RMSE) + 2n \qquad (5.18)$$

$$BIC = m\ln(RMSE) + n\ln(m) \qquad (5.19)$$

Where, m is the number of input–output patterns used for training, n is the number of parameters to be identified and RMSE is the root-mean-square error between the network output and target. The performance measures generally improves as more parameters are added to the model, but the AIC and BIC statistics penalize the model for having more parameters and, therefore, tend to result in more parsimonious models.

5.6.7 Uncertainty in Model Output

The predictive uncertainty of the models is evaluated by an index called the 'noise-to-signal ratio'. It is usually compared with the standard deviation of the observed values. The ratio of Standard Error of Estimate (SEE) to Standard Deviation of observed values, called the noise-to-signal ratio, indicates the degree to which noise hides the information. The unbiased SEE is a measure of the unexplained variance (Tokar and Johnson, 1999). If the SEE is significantly smaller than the standard deviation, then the model can provide accurate predictions. On the contrary, if the ratio is greater than or equal to unity, then the model predictions will not be accurate (McCuen, 1993).

$$SEE = \left[\frac{1}{\upsilon} \sum_{i=1}^{N} (Q_c - Q_o)^2 \right]^{0.5} \tag{5.20}$$

where Q_o, Q_c are the observed and computed values of flow N is the total number of observations, and υ is the degrees of freedom and is equal to the number of observations in the training set minus the number of parameters.

5.6.8 Optimal Model Parameter

The structure of either model is to be decided after taking a lot of hit and trial on deciding optimal parameters. In case of MLP models the influence on performances has been tested by varying the number of neurons to get a maximum performance. The parameter 'spread' is varied in case of RBF model and cluster 'radius' is varied in case of TS model.

5.7 RESULTS AND DISCUSSION

The DOWR is presently using time lag method in making the forecast. A standard time difference is being observed when a discharge passes from Khairmal to Barmul and Mundali. The locations between these stations and the travel time of discharges are given in Fig.5.6. A maximum of 24-32 hour is required by a flood peak to reach from Khairmal to Mundali.

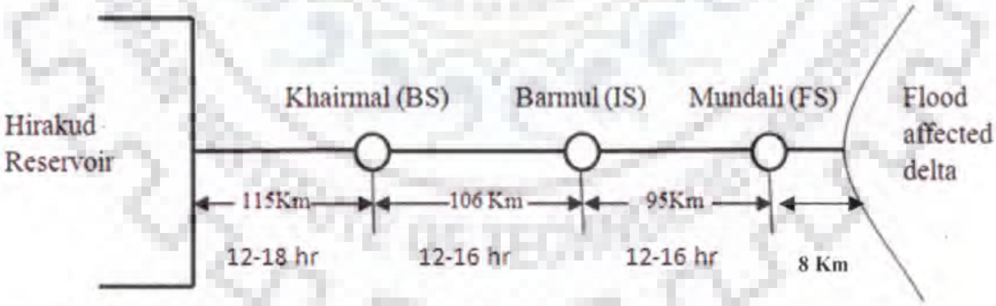


Fig. 5.6 Schematic diagram of B.S, I.S. and F.S with distance and travel time

The flood forecasting has been carried out considering the peak as well as 3-hour discharges as input vectors in the respective models. The results are given in following sections:

5.7.1 Forecasting based on peak discharges

3-hourly discharge data for a period from 1996-2010 at Khairmal, Barmul and Mundali have been collected from Department of Water Resources, Government of Orissa. The peak discharges at each station were derived from this data. For each G&D site 101 flood peaks of different magnitudes are obtained and out of these initial 50 peaks are considered for calibration of the all models and rest 51 for validation. The corresponding travel times of each peak is derived from the time series data of discharges of base and forecasting stations.

Statistical approach (LTF), MLFF-ANN, RBF-ANN and TS-fuzzy models are applied and compared using various model performance measuring criteria. During the flood time sometimes data of all the gauging sites are not available due to various reasons. Therefore, there may arise 3 cases during flood time for forecasting the discharge at Mundali (M). The tentative model combinations formed out of these cases are:

Case-I, When data of Khairmal (K) and Barmul (B) are available, relationship between KBM is developed.

KBM => $f(Q_M) = f(Q_K, Q_B)$ (5.21)

Case-II, When data of only Khairmal is available, one relation KM is developed and another relationship is developed to compute Mundali flow Khairmal and Bamul computed flows (Bc) for prediction at Mundali (KB_CM).

KM => $f(Q_M) = f(Q_K)$ (5.22)

KBcM => $f(Q_M) = f(Q_K, Q_{BC})$ (5.23)

Case-III, When data of only Barmul is available relationship BM is developed.

BM=> $f(Q_M) = f(Q_B)$ (5.24)

By using statistical approach following relationships are developed (Table 5.1). Simultaneously, two ANN and Fuzzy approaches are also applied on peak floods of above stations.

Table 5.1 Relationship between discharges using simple statistical method (peak to peak)

Combinations	Statistical relationships	R	
		Calibration	Validation
KB	$Q_B = 1.118Q_K - 206.78$	0.902	0.899
BM	$Q_M = 1.05Q_B + 382.19$	0.915	0.912
KM	$Q_M = 1.134Q_K + 11.507$	0.906	0.914
KBM	$Q_M = 0.834Q_B + 0.201Q_K + 184.03$	0.930	0.915
KBcM	$Q_M = 0.393Q_{BC} + 0.885Q_K + 34.2$	0.836	0.702

Where, Q_M = peak discharge at Mundali (Forecasting Station), Q_B = peak discharge at Barmul (Intermediate Station), Q_K = peak discharge at Khairmal (Base Station) and Q_{BC} = computed peak discharge at Barmul.

The same input vectors are again put into MLFF-ANN architecture using MATLAB codes. The trial has been taken with a 3-layered feed forward network. Different combinations of feed forward network have been attempted by changing transfer function, varying neurons from 1 to 7 and epochs from 50 to 300 at an increment of 25. The model performances of different feed forward network are noted. In all cases the hyperbolic tangent sigmoid transfer function 'tansig' are used in first layer, the linear transfer function 'purelin' in second layer and Bayesian regularization backpropagation 'trainbr' remains the network training function. Both 'tansig' and 'purelin' are better compatible with feed forward networks. The 'trainbr' training function updates the weight and bias values according to Levenberg-Marquardt optimization and Bayesian regularization process to produce a better response. The details of the optimal network structures applied for different combinations are shown in Table 5.2.

In the same way the combinations with the peak values are also put in RBF network. The parameter spread varied from 0.5-1.5 (Table 5.2). The RBF model almost gave a standard result at around 100 epochs.

For TS models, subtractive cluster radii varied from 0.1 to 0.8 with an increment of 0.05 in each trial. The fuzzy rules varied from 5 to 3 to get the optimal performance (Table 5.2).

Table 5.2 Optimal model parameters (peak flood)

Between stations	ANN			Fuzzy	
	MLFF	RBF		TS	
	Structure	Structure	Spread	Radius	Rules
KB	1-2-1	1-2-1	1.0	0.6	3
KBM	2-3-1	2-3-1	1.2	0.35	5
KM	1-2-1	1-2-1	0.8	0.65	3
BM	1-2-1	1-2-1	1.0	0.65	3
KBcM	2-3-1	2-3-1	1.2	0.35	5

The optimal model parameters are derived after lot of hit and trials. The influence of variation of neurons in the performance of individual MLFF models is tested. For convenience, variation of RMSE and Efficiency with number of neurons is shown in Fig.

5.7(a) to Fig. 5.8(d). It is clearly seen that the KM, BM combinations show minimum value at 2 neurons whereas KBM and KBcM model uses 3 neurons to achieve minimum RMSE. The same performance is also repeated in case of efficiency. Further, adding of neurons reduces the performances instead of improving it. Thus the optimal model structure for KM and BM is 1-2-1 and for KBM and KBcM, it is 2-3-1. The optimal model parameters for RBF model is also finalized by varying the spread parameter. It was varied from 0.5 to 1.5 with an increment of 0.1 to achieve the best performance. The variation of RMSE and Efficiency with spread is shown in Fig. 5.9(a) to 5.10(d). The optimal neuron structure as obtained from RBF is 1-2-1 for KM, BM and KB combination and 2-3-1 for KBM and KBcM. By varying the cluster radius in case of TS model the changes have been observed in model performances. The radius was varied with an increment of 0.05 in each attempt. The number of rules gets varied with varying cluster radius. The variations in RMSE and Efficiency with cluster radius are shown in Fig. 5.11(a) to 5.12 (d).

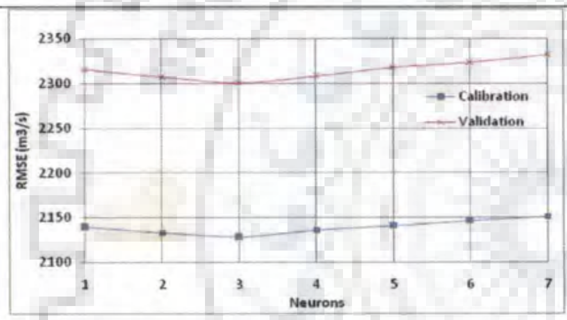


Fig. 5.7(a) Variation of RMSE with number of hidden neurons in KBM combination

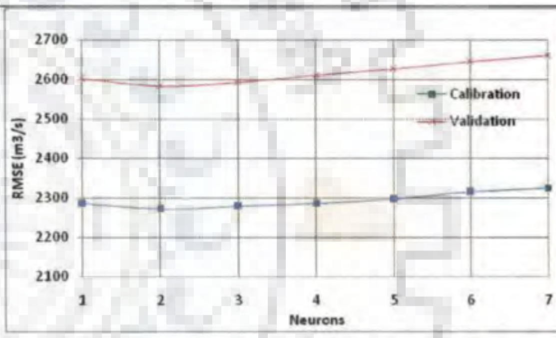


Fig. 5.7(b) Variation of RMSE with number of hidden neurons in KM combination

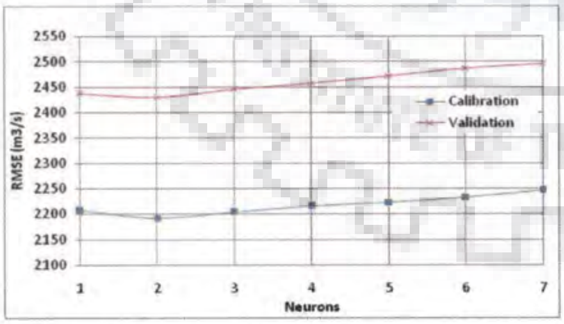


Fig. 5.7(c) Variation of RMSE with number of hidden neurons in BM combination

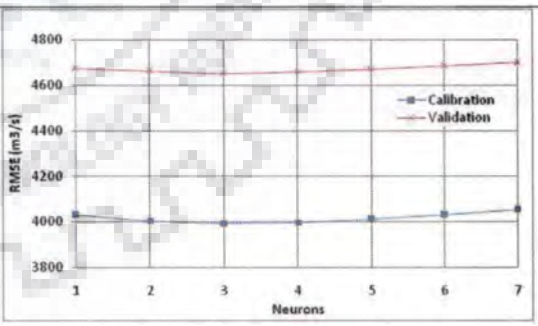


Fig. 5.7(d) Variation of RMSE with number of hidden neurons in KBcM combination

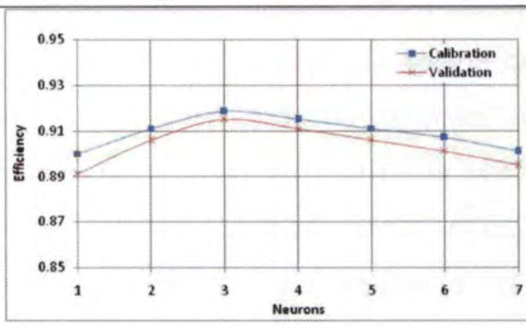


Fig. 5.8(a) Variation of efficiency with number of hidden neurons in KBM combination

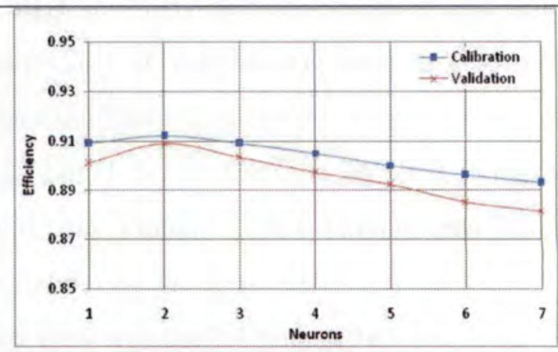


Fig. 5.8(b) Variation of efficiency with number of hidden neurons in KM combination

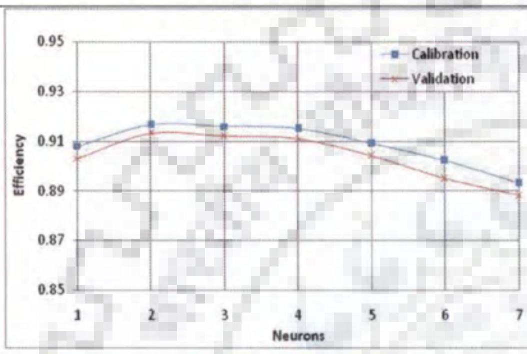


Fig. 5.8(c) Variation of efficiency with number of hidden neurons in BM combination

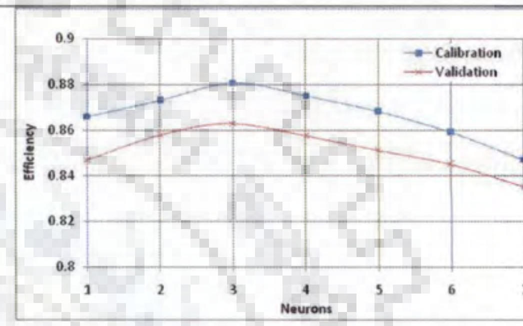


Fig. 5.8(d) Variation of efficiency with number of hidden neurons in KBcM combination

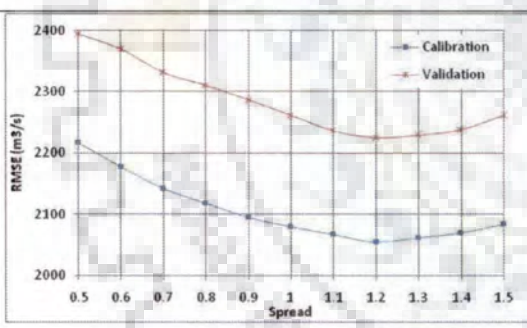


Fig. 5.9(a) Variation of RMSE with Spread in KBM

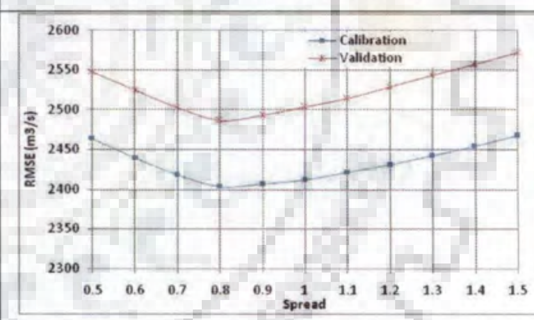


Fig. 5.9(b) Variation of RMSE with Spread in KM

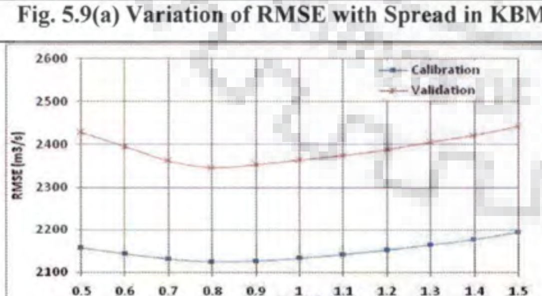


Fig. 5.9(c) Variation of RMSE with Spread in BM

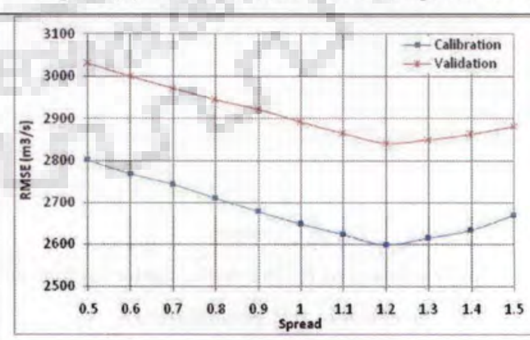


Fig. 5.9(d) Variation of RMSE with Spread in KBcM

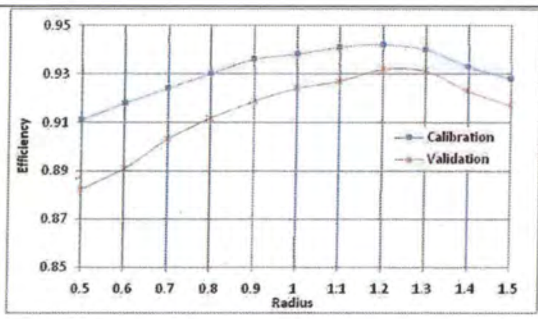


Fig. 5.10(a) Variation of efficiency with Spread in KBM

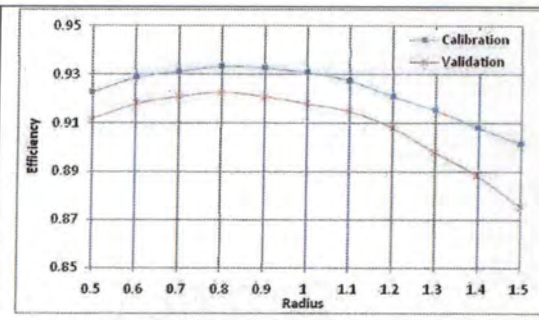


Fig. 5.10(b) Variation of efficiency with Spread in KM

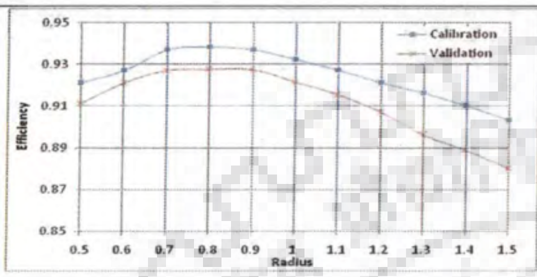


Fig. 5.10(c) Variation of efficiency with Spread in BM

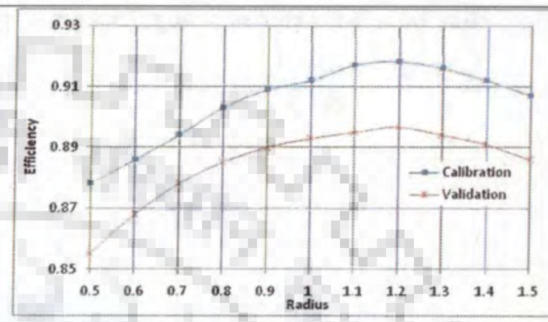


Fig. 5.10(d) Variation of efficiency with Spread in KBcM

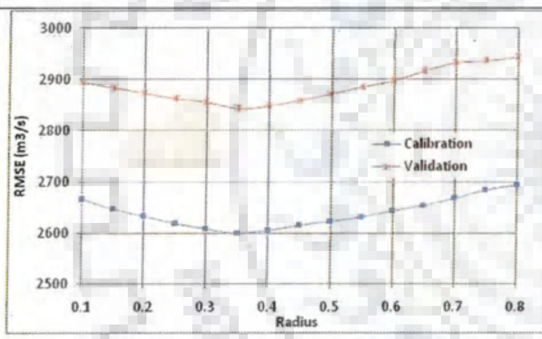


Fig. 5.11(a) Variation of RMSE with Radius in KBM

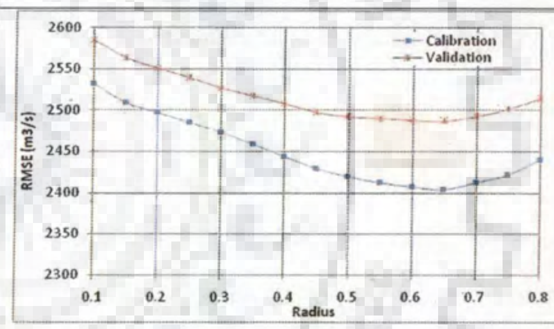


Fig. 5.11(b) Variation of RMSE with Radius in KM

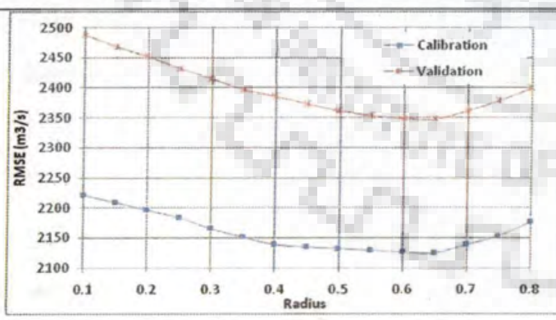


Fig. 5.11(c) Variation of RMSE with Radius in BM

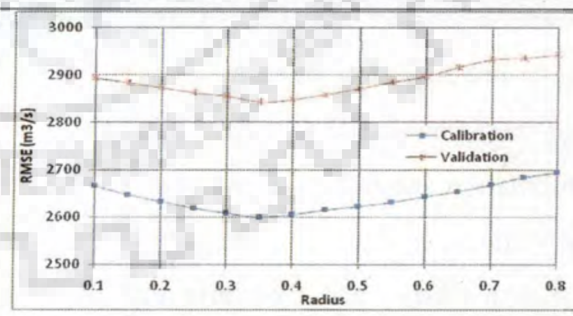


Fig. 5.11(d) Variation of RMSE with Radius in KBcM

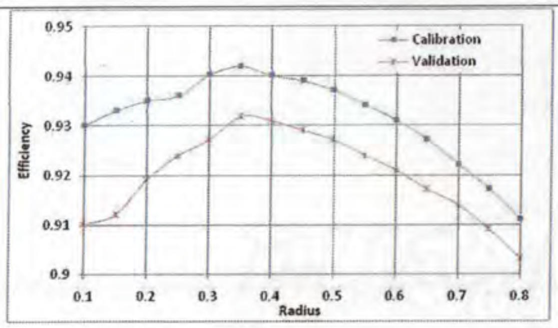


Fig. 5.12(a) Variation of efficiency with Radius in KBM

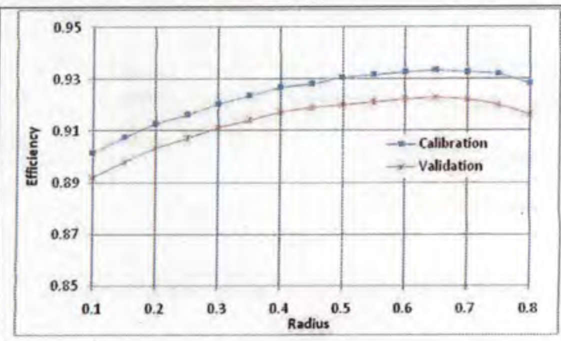


Fig. 5.12(b) Variation of efficiency with Radius in KM

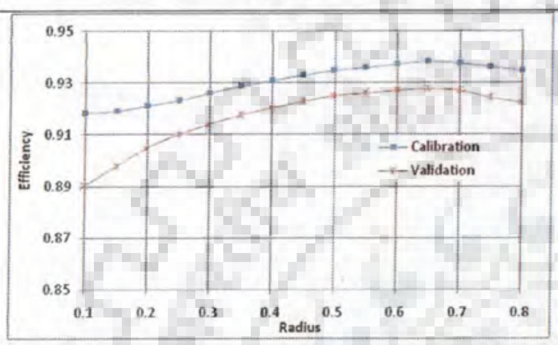


Fig. 5.12(c) Variation of efficiency with Radius in BM

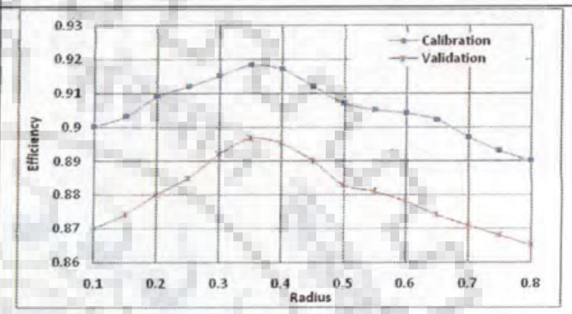


Fig. 5.12(d) Variation of efficiency with Radius in KBcM

The performance of all models for all four combinations like KBM, KM, BM and KBcM are given in Table 5.3. The KBcM combination is formed with peak discharge at Khairmal and computed value of Barmul (from KB combination) to derive the peak discharge at Mundali. It is revealed from Table 5.3 that in all cases the TS model has shown better performance as compared to RBF, MLP and LTF models. RBF has also shown better performance as compared to MLP and LTF. The LTF model has shown poor performance than other models in all cases. In combination wise the KBM combination shows better performance than KM and BM. But it requires information of two stations; in that case KM may be a better option as it saves very important warning time with a little loss of performance. The performance of BM although better than KM, for that hardly a little time (say 12 hours) will be left for taking relief initiatives. But the KBcM combination utilizing the computed value at Barmul from KB combination would have been a better option if its performance excels that of KM. But only an efficiency of 89.67% and RMSE of 2842.6 m^3/s in validation is achieved in case of TS model. Results of other models for this combination are still poorer than this. Thus, it can be concluded that performance wise KBM combination is better but KM combination with a little sacrifice in performance saves precious warning time.

Table 5.3 Performance criteria of all models (peak flood)

		TS		RBF		MLP		LTF	
		Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation
KBM	R	0.9809	0.9782	0.9775	0.9684	0.9742	0.9674	0.93	0.915
	RMSE(m3/s)	2054.5	2224.6	2091.3	2251.3	2128.1	2300.4	2399.3	2516.2
	Eff.	0.942	0.9318	0.9351	0.9244	0.9185	0.9151	0.911	0.901
KM	R	0.9712	0.9623	0.9702	0.9653	0.9705	0.9651	0.906	0.914
	RMSE(m3/s)	2403.2	2486.4	2446.1	2539.5	2272.1	2582.6	3562.4	3043.1
	Eff.	0.9333	0.9228	0.9212	0.9203	0.9119	0.9091	0.885	0.862
BM	R	0.98	0.9792	0.9773	0.9654	0.9753	0.9652	0.915	0.912
	RMSE(m3/s)	2123.9	2346.8	2141.6	2407.8	2192	2430.7	3156.78	3050.1
	Eff.	0.9381	0.9275	0.9367	0.9233	0.9166	0.9131	0.901	0.87
KBcM	R	0.9624	0.9538	0.9605	0.9451	0.951	0.9444	0.836	0.702
	RMSE(m3/s)	2598.8	2842.6	2633.1	2926.6	2683.5	2981.9	3992.9	4651.75
	Eff.	0.9182	0.8967	0.9119	0.8828	0.8808	0.8629	0.801	0.69

The scatter plot is made for the four combinations KBM, KM, BM and KBcM for all models. It is seen that statistical models gave under value for almost all combinations but the KBcM values are mostly erratic (Fig. 5.13(a)). Low and medium peaks gave somewhat better values for all combination but the high peaks are very poorly modeled in all combinations.

In the scatter plot of TS model (Fig. 5.13(b)) the KM, KBM models show good results for medium and high peaks although the low peaks remain somewhat under-predicted. The KBcM model unable to predict medium and low peaks closely and in most cases it remains under predicted. Similarly results are also obtained from RBF and MLP scatter plot performances. RBF model remains over predicted for high peaks and MLP shows under prediction. However, result of TS scatter plot is more uniform than that of RBF and MLP (Fig. 5.13(c) and 5.13(d)).

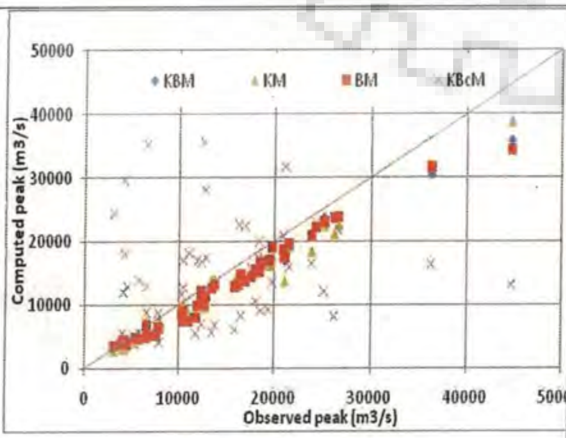


Fig. 5.13(a) Scatter plot of all 4 combinations using statistical model

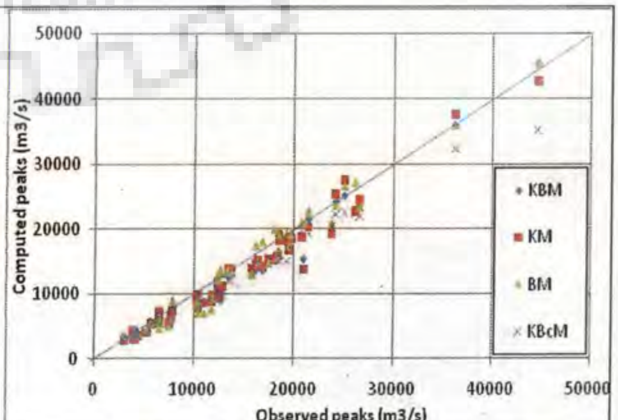


Fig. 5.13(b) Scatter plot of all 4 combinations using Fuzzy (TS) model

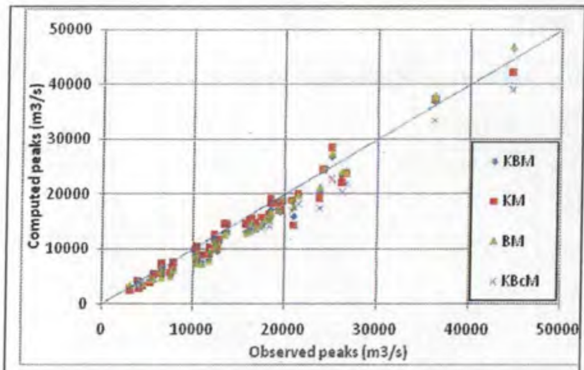


Fig. 5.13(c) Scatter plot of all 4 combinations using RBF model

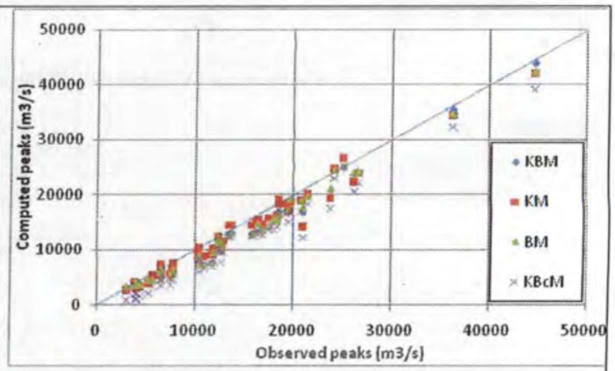


Fig. 5.13(d) Scatter plot of all 4 combinations using MLP model

5.7.1.1 Computation of Travel Times

For travel time the corresponding discharge and travel times between the stations are trialed with the same calibration and testing data range applying both statistical, ANN and fuzzy architectures. But none of the case has shown a better efficiency. So to decide travel times, clustering methods are adopted for the discharge value at base station and its corresponding travel time for the calibration dataset. At least 5 clusters are attempted applying K-mean and Fuzzy c-mean method using the MATLAB code. The datasets are divided into 5 clusters and their varying ranges are recorded. For each range the average travel time has been calculated along with average peak values at Khairmal. Basing on these average values a plot between peak discharge and corresponding travel times are drawn. Similarly the travel times also plotted for Barmul to Mundali discharge (Fig. 5.14 (a)). While using peak flood forecasting this plot may be used as a reference for finding the corresponding travel time of the forecasted peak.

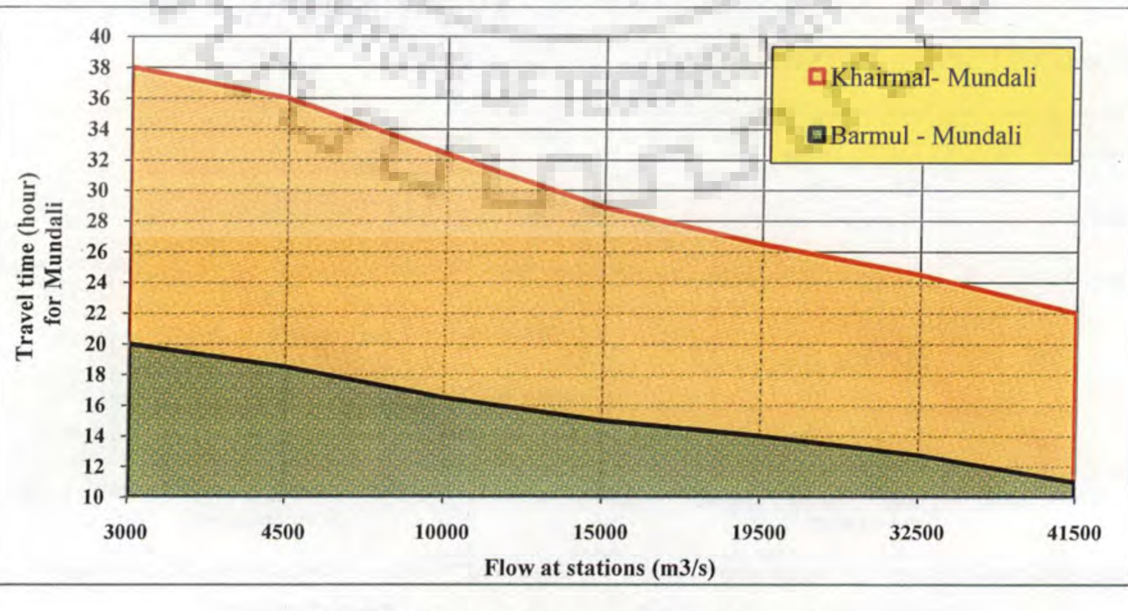


Fig. 5.14(a) Travel times against the peaks

5.7.1.2 Discussion on Peak Flood Forecasting

Statistical, ANN and Fuzzy methods are applied for downstream Hirakud catchment between stations Khairmal, Barmul and Mundali. As per the model performance criteria TS-fuzzy method has shown a better performance than MLFF-ANN, RBF-ANN and LTF methods (Table 5.3). Whereas,performance of RBF has a little improvement over MLFF. The statistical method (LTF) has not performed satisfactorily.

For selection of best combination of stations, here 3 combinations directly (KM, BM, KBM) and one derived combination (KBcM) was attempted. It was realized that the KBM combination gives the better results followed by BM and KM. The performance of derived combination KBcM has been observed as poorer to KM and therefore it is not recommended. The performance wise although KBM is better but simultaneously KM combination can also be recommended as it saves precious lead time. The RMSE and efficiency of TS model for KBM combination is 2224.6, 0.9318 and that of KM is 2486.4, 0.9228 in validation. The average travel time of flood from Khairmal to Barmul is about 12 hours. So, for a high peak at least 12 hour is saved if KM combination is used instead of KBM. The BM combination may be used if KM information is not available or may be used as a cross check. Table 5.4 provides the best models for forecasting the floods at Mundali. For a quick reference of predicted peaks (on the result of TS method) at Mundali with inputs of observed peaks at either Khairmal or Barmul is shown in Fig. 5.14(b). On the basis of the availability of the data the users can directly select a suitable model with optimal parameter to ascertain the defined efficiency.

Table 5.4 Best Flood forecasting models at Mundali gauging site

Model combination	**Data available**		**Forecast at Mundali using**	**TS Model Efficiency**
	Khairmal	**Barmul**		
Case-I	Yes	Yes	KBM=> $f(Q_M) = f(Q_K, Q_B)$	0.9318
Case-II	Yes	No	KM=> $f(Q_M) = f(Q_K)$ KBcM=> $f(Q_M) = f(Q_K, Q_{BC})$	0.9228 0.8967
Case-III	No	Yes	BM=> $f(Q_M) = f(Q_B)$	0.9275

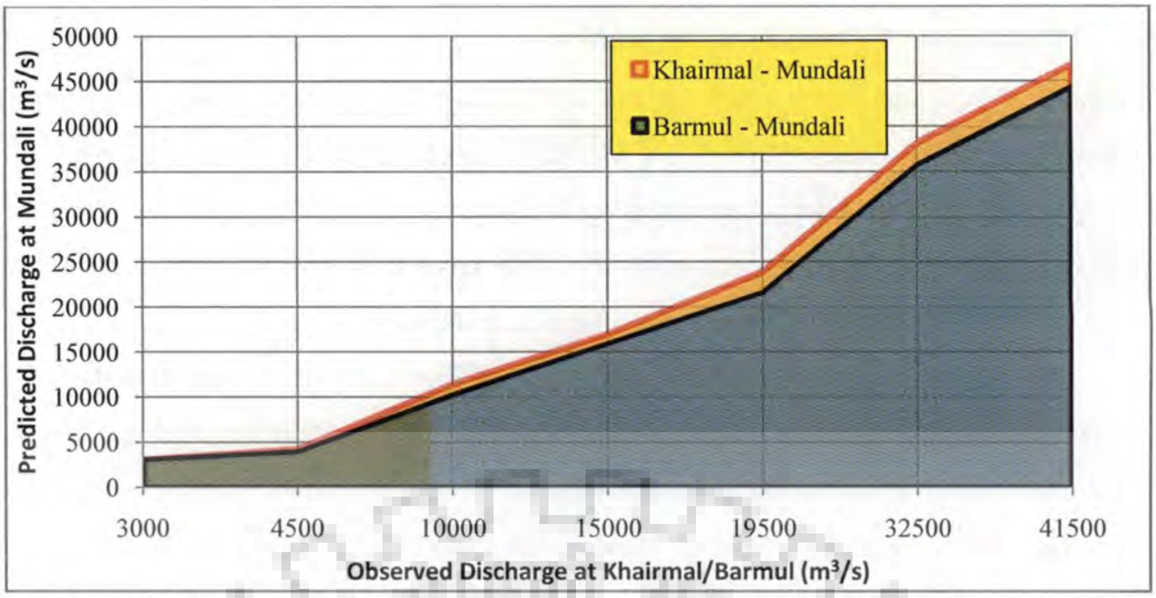


Fig. 5.14(b) Forecasted discharge at Mundali against the Observed peaks of either Khairmal or Barmul (TS method)

5.7.2 FORECASTING BASED ON 3-HOURLY DISCHARGES

The 3-hourly discharges as collected from base station at Khairmal, intermediate station at Barmul and Forecasting station at Mundali for the period of 1996-2010 are used for the forecasting. The statistical method based on linear transfer function (LTF), ANN methods based on MLFF and RBF network and fuzzy method based on Takagi Sugeno (TS) approach have been applied in this study. Different models with separate combination of inputs were selected and intercompared.

5.7.2.1 Model development

One of the most important steps in establishment of model is the determination of significant input vectors. The statistical parameters such as auto correlation function (ACF), partial auto correlation function (PACF) and cross correlation function (CCF) are used for this purpose. The pictorial representation of cross correlation between Khairmal with Mundali and Barmul with Mundali is given in Fig. 5.15(a) and 5.15(b), from where the maximum lags are derived. The ACF series of Khairmal and Barmul G&D stations reveals that the individual stages are autoregressive (Fig. 5.16(a) and 5.16(b)). The respective PACF values of the flow with 95% confidence level gives potential antecedent flow values that have influence on the flow value at the current period. When comparing with the flow data of forecasting station (Mundali), the base station Khairmal show a higher CCF at Lag 9 and station Barmul at Lag 5 (Fig.5.17). In all cases one lag is equivalent to 3 hours as frequency of the discharge data is 3 hours.

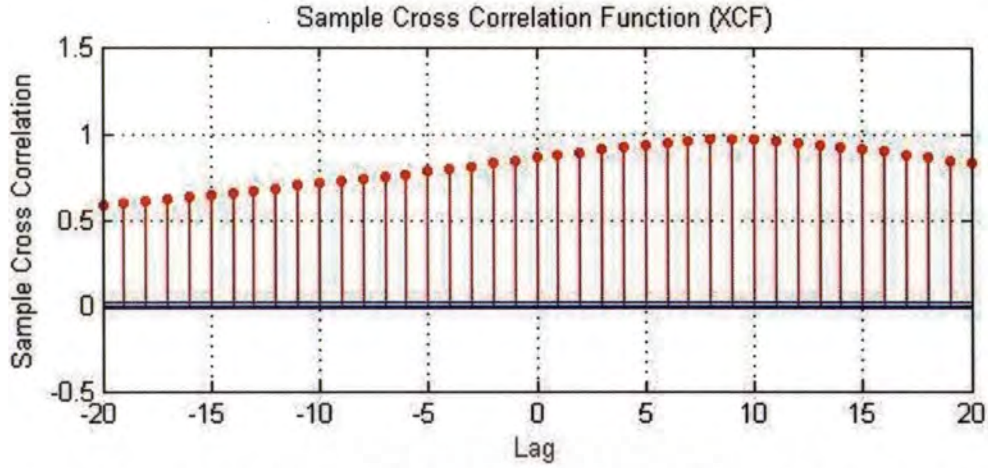


Fig. 5.15(a) Pictorial information of lag correlations between Khairmal and Mundali

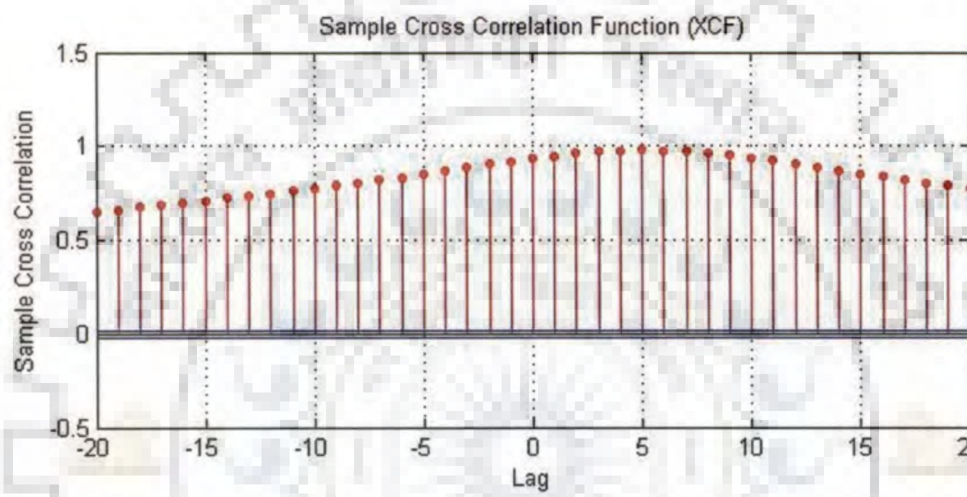


Fig. 5.15(b) Pictorial information of lag correlations between Barmul and Mundali

From PACF figures (Fig. 5.16(a) and 5.16(b)) it is observed that a lag upto 9 may be included in Khairmal and upto 5 lags for Barmul. Therefore, on the basis of PACF and CCF of the data series, the input vectors have been selected for flood forecasting models in the present study.

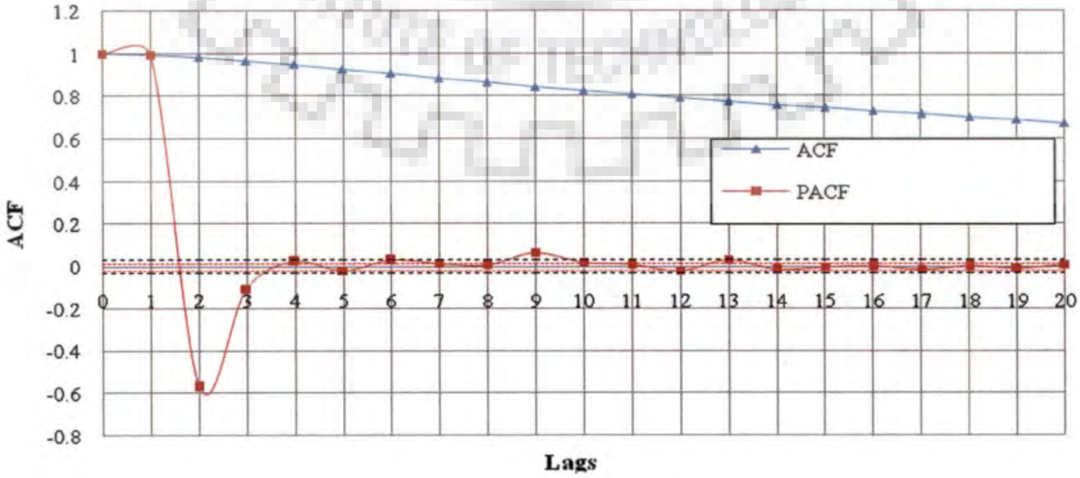


Fig. 5.16(a) Auto-correlation and partial auto-correlation of Khairmal discharge

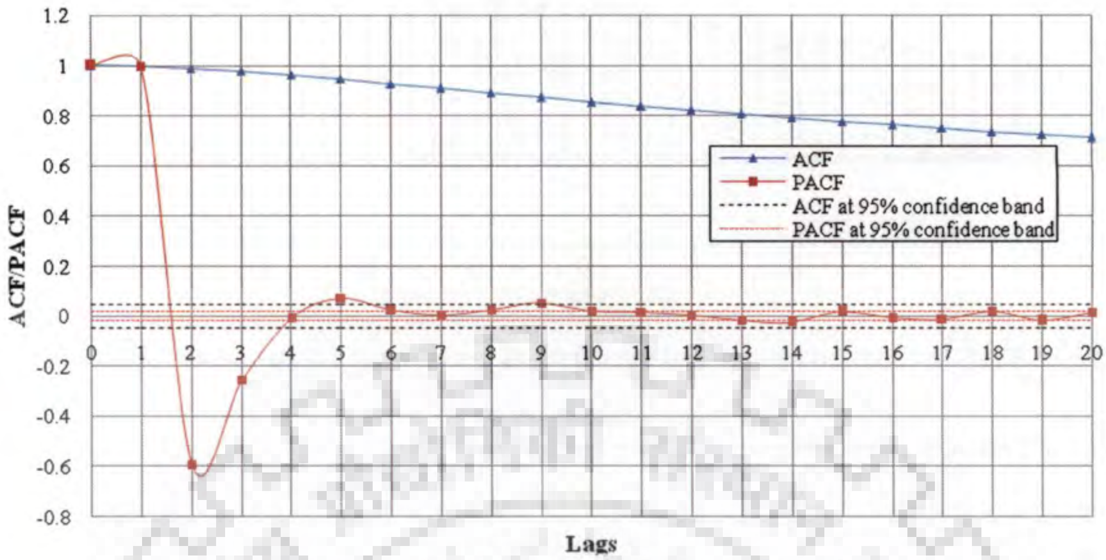


Fig. 5.16(b) Auto-correlation and partial auto-correlation of Barmul discharge

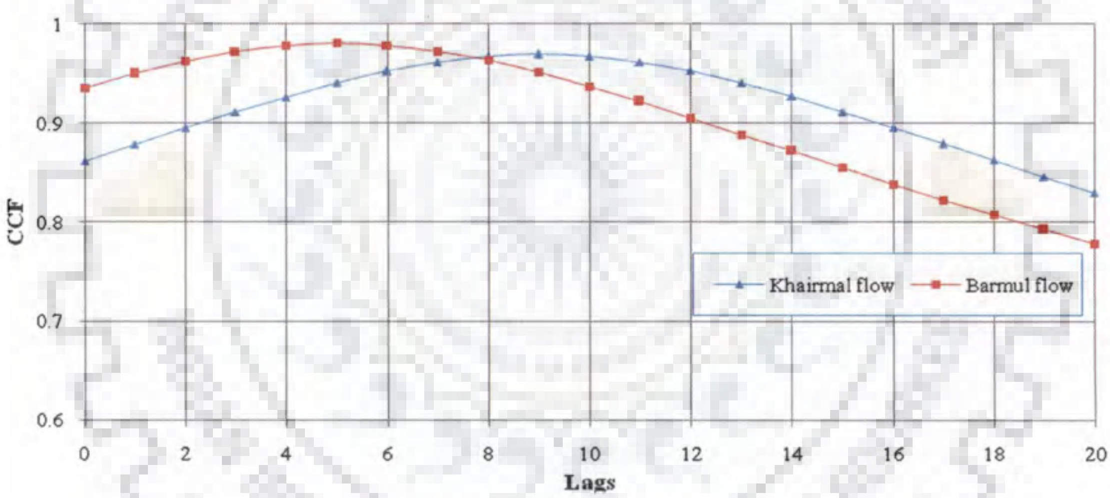


Fig. 5.17 Cross correlation of discharge at Mundali with discharge at Khairmal and Barmul

The formation of the possible models is decided on the basis of best lag instead of starting trial with $Q_{K(t-1)}$ and so on. Khairmal discharge had shown a good correlation with discharge of Mundali at 9th lag. So the first model combination is kept as $Q_{M(t)} = f(Q_{K(t-9)})$). Keeping the 9th lag at center, other inputs like $Q_{K(t-10)}, Q_{K(t-8)}$ are added to first combination and subsequent combinations are formed gradually. In this way 9 combinations are formed taking $Q_{K(t-13)}$ to $Q_{K(t-5)}$ using the information of Khairmal only. Similarly, using the flow information of Barmul only and keeping the maximum lag of Barmul $Q_{B(t-5)}$ at center five more combinations are formed and trialed.

The combine model combination is planned to be formed by joining both the best combinations with Khairmal and Barmul data respectively. On the basis of various input vectors a number of combinations have been formed as shown in Table 5.5 to be put on different models for testing their performances.

Table 5.5 Formation of forecasting models with different input vector combinations

Case I: Input vector combinations with flow data of Khairmal only	
KM-1	$Q_{M(t)} = f(Q_{K(t-9)})$
KM-2	$Q_{M(t)} = f(Q_{K(t-9)}, Q_{K(t-8)})$
KM-3	$Q_{M(t)} = f(Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)})$
KM-4	$Q_{M(t)} = f(Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)})$
KM-5	$Q_{M(t)} = f(Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)})$
KM-6	$Q_{M(t)} = f(Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)})$
KM-7	$Q_{M(t)} = f(Q_{K(t-12)}, Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)})$
KM-8	$Q_{M(t)} = f(Q_{K(t-12)}, Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)}, Q_{K(t-5)})$
KM-9	$Q_{M(t)} = f(Q_{K(t-13)}, Q_{K(t-12)}, Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)}, Q_{K(t-5)})$
Case II: Input vector combinations with flow data of Barmul only	
BM-1	$Q_{M(t)} = f(Q_{B(t-5)})$
BM-2	$Q_{M(t)} = f(Q_{B(t-5)}, Q_{B(t-4)})$
BM-3	$Q_{M(t)} = f(Q_{B(t-6)}, Q_{B(t-5)}, Q_{B(t-4)})$
BM-4	$Q_{M(t)} = f(Q_{B(t-6)}, Q_{B(t-5)}, Q_{B(t-4)}, Q_{B(t-3)})$
BM-5	$Q_{M(t)} = f(Q_{B(t-7)}, Q_{B(t-6)}, Q_{B(t-5)}, Q_{B(t-4)}, Q_{B(t-3)})$
Case III: Input vector combinations with flow data of Khairmal and Barmul combined	
KBM	Combination of best model of Khairmal and Barmul

Forecasting at Barmul using Khairmal information and then using this computed discharge for forecasting at Mundali has given a poor performance than the Khairmal-Mundali combination. The similar results have also been found while forecasting the peak discharge at Mundali using computed discharge of Barmul. So, to avoid complexity the combination (KBcM) is not considered in further study.

5.7.2.2 LTF model

Different linear transfer functions models have been developed using the various combinations of three different cases given in Table 5.5. In Case-I, the model efficiency varies from 0.8825 to 0.8939 during calibration and from 0.8531 to 0.8717 during validation (Table 5.6). The root mean square value varies from 1700.20 to1884.36 m^3/s during calibration and from 1939.10 to 2178.14m^3/s during validation. Similarly, the coefficient of correlation varies from 0.9235 to 0.9331 during calibration and 0.8905 to 0.9081 during validation. In Case-II, the model efficiency varies from 0.8935 to 0.9010 during calibration and from 0.868 to 0.8791 during validation. The root mean square value varies from 1683.00 to 1761.00m^3/s during calibration and from 1906.00 to 1959.00m^3/s during validation. Similarly, the coefficient of correlation varies from 0.9240 to 0.9292 during calibration and 0.8820 to 0.8891 during validation.

5.7.2.3 ANN model

The flood forecasting models have also been developed through Multi-layered feed forward and RBF networks as per the input vectors of Table 5.5.

(A) MLP Model

The feed forward hierarchical architecture is the most commonly used neural network structure (Maier and Dandy, 2000). The inputs are tried with 1 to 12 neurons in hidden layer. The epochs are taken from 50 to 500 with 25 increments in each iteration. The 'trainbr' function updates the weight and bias values as per Levenberg-Marquardt optimization. Other parameters are kept same as for discrete values. The performance criteria of each network remains fixed as R, RMSE and Efficiency.

The MLP models developed using nine different input vectors in Case-I indicate that the model efficiency varies from 0.9084 to 0.9217 during calibration and from 0.8998 to 0.9116 during validation (Table 5.6). The root mean square value varies from 1438.70 to 1376.80 m^3/s during calibration and from 1506.30 to 1721.10 m^3/s during validation. Similarly, the coefficient of correlation varies from 0.9518 to 0.9629 during calibration and 0.9325 to 0.9439 during validation. In Case-II, the MLP model efficiency varies from 0.9266 to 0.9316 during calibration and from 0.9094 to 0.9147 during validation. The root mean square value varies from 1358.50 to 1416.10 m^3/s during calibration and from 1486.10 to 1566.20 m^3/s during validation. Similarly, the coefficient of correlation varies from 0.9563 to 0.9671 during calibration and 0.9503 to 0.9611 during validation.

(B) RBF Model

Once the input vector is decided, the OLS algorithm optimizes the RBF network architecture; however, for an MLP, an arbitrarily large amount of network with varying hidden units needs to be constructed and a best-fit network is decided based on performance criteria. The radial basis network is designed considering the goal and spread. Since the hidden neurons of the RBF networks are fixed by the OLS algorithm, the usual trial-and-error procedure for identifying the appropriate network structure is eliminated. In this study, the goal varied from 0.5 to 1.2 and spread varied from 1.0 to 2.2. Very small incremental change has been initiated in each trial. Very minute change has been observed by changing the goals but a significant change has been visualized by varying the spread of the RBF architecture.

The RBF model efficiency for the 9 different models of Case-I varies from 0.9151 to 0.9262 during calibration and from 0.9010 to 0.9139 during validation (Table 5.6). The root mean square value varies from 1369.30 to 1416.60 m^3/s during calibration and from 1496.30 to 1687.80 m^3/s during validation. Similarly, the coefficient of correlation varies from 0.9547 to 0.9630 during calibration and 0.9350 to 0.9495 during validation. In Case-II, the RBF model efficiency varies from 0.9186 to 0.9320 during calibration and from 0.9098 to 0.9169 during validation. The root mean square value varies from 1357.20 to 1437.20 m^3/s during calibration and from 1473.00 to 1552.5.00 m^3/s during validation. Similarly, the coefficient of correlation varies from 0.9578 to 0.9687 during calibration and 0.9574 to 0.9610 during validation.

5.7.2.4 Fuzzy (TS) model

In this study, the fuzzy membership functions and rules have been developed in such a way that the RMSE and other performance criteria between the observed and the predicted set of values are minimized. The adjustment of the membership functions and rules is accomplished by using the subtractive clustering analysis. The fuzzy logic based flood forecasting model has been developed in two steps:

(i) In this study the subtractive cluster radius varies from 0.1 to 0.7 with an increment of 0.05 in each trial. For each set of inputs different number of fuzzy rules are attempted depending on the subtractive radius fixed, and

(ii) optimization based on linear least square estimation

In Case-I, The TS model efficiency varies from 0.9214 to 0.9347 during calibration and from 0.9086 to 0.9230 during validation. The root mean square value varies from 1334.90 to 1405.20 m^3/s during calibration and from 1402.50 to 1560.10m^3/s during validation (Table 5.6). Similarly, the coefficient of correlation varies from 0.9682 to 0.9769 during calibration and 0.9567 to 0.9645 during validation. In Case-II, the TS model efficiency varies from 0.9334 to 0.9366 during calibration and from 0.9230 to 0.9283 during validation. The root mean square value varies from 1319.00 to 1407.00m^3/s during calibration and from 1371.00 to 1429.10m^3/s during validation. Similarly, the coefficient of correlation varies from 0.9586 to 0.9685 during calibration and 0.9621 to 0.9660 during validation.

Table 5.6 Performance of models with either Khairmal or Barmul flows only

Forecasting at **Mundali** with information of **Khairmal**								
Model	**KM-1**	$Q_{M(t)} = f(Q_{K(t-9)})$						
Performance measures	MLFF		RBF		TS		LTF	
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation
R	0.9518	0.9325	0.9547	0.935	0.9682	0.9567	0.9235	0.8905
RMSE	1438.7	1721.1	1416.6	1687.8	1405.2	1560.1	1884.36	2178.14
Efficiency	0.9084	0.8998	0.9151	0.901	0.9214	0.9086	0.8825	0.8531
KM-2	$Q_{M(t)} = f(Q_{K(t-9)}, Q_{K(t-8)}))$							
R	0.956	0.936	0.9572	0.9391	0.9716	0.9602	0.9270	0.8930
RMSE	1482.9	1680.8	1471.8	1629.6	1472.3	1518.9	1837.57	2104.17
Efficiency	0.9112	0.9021	0.9169	0.9038	0.9241	0.912	0.8866	0.8564
KM-3	$Q_{M(t)} = f(Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)})$							
R	0.9582	0.939	0.9591	0.9432	0.9729	0.9606	0.9298	0.8945
RMSE	1454.8	1647.3	1442.5	1595.6	1441	1509.1	1798.04	2052.32
Efficiency	0.9129	0.9042	0.9186	0.9052	0.9276	0.9139	0.8891	0.8594
KM-4	$Q_{M(t)} = f(Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)})$							
R	0.9602	0.9399	0.961	0.944	0.9746	0.9634	0.9309	0.8958
RMSE	1428.7	1582.4	1420	1567.6	1398.6	1448.5	1766.82	2012.97
Efficiency	0.9148	0.9061	0.9212	0.9081	0.9309	0.9161	0.8903	0.8615
KM-5	$Q_{M(t)} = f(Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)})$							
R	0.961	0.9408	0.9617	0.9447	0.9756	0.9637	0.9316	0.9068
RMSE	1411.9	1556.5	1405.6	1532.2	1374.9	1441.2	1737.9	1974.0
Efficiency	0.9168	0.9085	0.9227	0.9096	0.9317	0.9197	0.8918	0.8638
KM-6	$Q_{M(t)} = f(Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)})$							
R	0.9621	0.942	0.9624	0.9473	0.9765	0.9642	0.9322	0.9075
RMSE	1394.2	1535.6	1371	1503.7	1351.5	1406.9	1716.8	1952.0
Efficiency	0.9204	0.91	0.9249	0.9109	0.9335	0.9225	0.8931	0.8685
KM-7	$Q_{M(t)} = f(Q_{K(t-12)}, Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)})$							
R	0.9626	0.9435	0.9628	0.9493	0.9768	0.9644	0.9329	0.9080
RMSE	1377.4	1507.1	1370.1	1497	1336.4	1403.4	1702.9	1941
Efficiency	0.9215	0.9114	0.9260	0.9136	0.9346	0.9228	0.8936	0.8716
KM-8	$Q_{M(t)} = f(Q_{K(t-12)}, Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)}, Q_{K(t-5)})$							
R	0.9628	0.9438	0.9629	0.9494	0.9769	0.9645	0.9330	0.9081
RMSE	1377.1	1506.8	1369.8	1496.7	1335.8	1403.2	1701.0	1940.2
Efficiency	0.9216	0.9114	0.9261	0.9138	0.9347	0.9229	0.8938	0.8717
KM-9	$Q_{M(t)} = f(Q_{K(t-13)}, Q_{K(t-12)}, Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)}, Q_{K(t-5)})$							
R	0.9629	0.9439	0.9630	0.9495	0.9769	0.9645	0.9331	0.9081
RMSE	1376.8	1506.3	1369.3	1496.3	1334.9	1402.5	1700.2	1939.1

Efficiency	0.9217	0.9116	0.9262	0.9139	0.9347	0.9230	0.8939	0.8717
	Forecasting at **Mundali** with information of **Barmul** and **Khairmal**							
BM-1	$Q_{M(t)} = f(Q_{B(t-5)})$							
R	0.9563	0.9503	0.9578	0.9574	0.9586	0.9621	0.9240	0.8820
RMSE	1416.8	1566.2	1437.2	1552.5	1407	1429.1	1761	1959
Efficiency	0.9266	0.9094	0.9186	0.9098	0.9334	0.9230	0.8935	0.8680
BM-2	$Q_{M(t)} = f(Q_{B(t-5)}, Q_{B(t-4)})$							
R	0.9596	0.9535	0.964	0.9607	0.9679	0.9651	0.9288	0.8870
RMSE	1384.8	1525.6	1379.7	1495.1	1359.9	1399.5	1704	1916
Efficiency	0.9293	0.9125	0.925	0.9136	0.9359	0.9274	0.8999	0.8736
BM-3	$Q_{M(t)} = f(Q_{B(t-6)}, Q_{B(t-5)}, Q_{B(t-4)})$							
R	0.9665	0.9606	0.9686	0.9609	0.9681	0.9655	0.9291	0.8890
RMSE	1362.4	1486.5	1360.2	1473.5	1321.3	1371.6	1686.0	1909.0
Efficiency	0.9314	0.9145	0.9320	0.9167	0.9364	0.9283	0.9010	0.8790
BM-4	$Q_{M(t)} = f(Q_{B(t-6)}, Q_{B(t-5)}, Q_{B(t-4)}, Q_{B(t-3)})$							
R	0.9667	0.9609	0.9687	0.9610	0.9684	0.9657	0.9292	0.8890
RMSE	1360.2	1486.3	1358.1	1473.2	1320.3	1371.3	1684.0	1907.0
Efficiency	0.9315	0.9146	0.9320	0.9168	0.9365	0.9282	0.9010	0.8791
BM-5	$Q_{M(t)} = f(Q_{B(t-7)}, Q_{B(t-6)}, Q_{B(t-5)}, Q_{B(t-4)}, Q_{B(t-3)})$							
R	0.9671	0.9611	0.9687	0.9610	0.9685	0.9660	0.9292	0.8891
RMSE	1358.5	1486.1	1357.2	1473.0	1319.0	1371.0	1683.0	1906.0
Efficiency	0.9316	0.9147	0.9320	0.9168	0.9366	0.9283	0.9010	0.8791

A combine model is also considered by combining any of the KM model (KM-1 to KM-9) with that of any of BM model (BM-1 to BM-5). It is seen that the combination KM-7 and BM-3 has given a improved performance than other combinations. So another input vector set KBM is formed which is the combination of KM-7 and BM-3. The performance of this KBM is also better than either of KM-7 and BM-3. The performance of KBM is kept as the base performance to find the parsimony of other input combinations which is discussed in next section.

5.7.2.5 Comparison of different models during calibration and validation

The performance of each combination as per Table 5.5 is recorded in Table 5.6. In general, a R^2 value greater than 0.9 (R>0.948) indicates a very satisfactory model performance, while a R^2 value in the range 0.8–0.9 (R=0.948 to 0.894) signifies a good performance and values less than 0.8 (R<0.894) indicate an unsatisfactory model performance (Coulibaly and Baldwin, 2005).

As far as these performances are concerned, it is seen that for Case-I, under MLP all models show the very satisfactory performance as the R values are greater than 0.948 in calibration where as in validation none of the values attain this range. In RBF all R values are above 0.948 in calibration but in validation the satisfactory result comes after

KM-7. All TS model values in calibration and validation show R value greater than 0.948. In LTF model the R value remain in satisfactory (R=0.894 to 0.948) range in all cases in calibration and in validation it begins from KM-3.

In Case-II, in all cases of MLP, RBF and TS the performance of R is satisfactory both in calibration and validation range. But in LTF the range remains in satisfactory level in calibration and validation.

In all Case I and II, RMSE value is around 3 to 5% of the maximum peak occurred in 2008 (44742m^3/s). Efficiency achieved in all cases is above 0.9 other than LTF. Therefore, in general all the models are providing the flood forecast in a very close range and can be considered independently depending on the availability of the input data and its suitability for use.

This becomes possible due to the input selection based on maximum lags as in case of starting with $Q_{k(t-1)}$ an efficiency of 0.81 and 0.78 is only achieved in calibration and validation.

In all cases the TS model extracts a better performance than MLFF, RBF or LTF models. The LTF models provide an inferior performance in comparison to others. The KM-9 model shows a high performance in all cases of MLFF, RBF, TS and LTF approaches when only Khairmal information is available. The performance of KM models starts increasing with addition of one more flow value on each attempt. The additional flow values are taken alternatively from both sides of lag 9 i.e (t-9). After reaching at KM-7 the performance increase becomes extremely slow. Although there is a very negligible increase in the performance, for that one more extra parameter gets added to the model combination. This addition of extra parameter might hamper the model parsimony. Thus all the models are tested for model parsimony in order to find the appropriateness of number of input vectors for producing the best performance. Proceeding in the same way for Barmul the BM-5 model performed best in all networks when only Barmul information is available. However, on the basis of AIC and BIC values it is confirmed that the model KM-7 and BM-3 and their combination model KBM show the parsimonious model. This KBM gives the most efficient performance as it covers all information of upstream and middle part of the river segment as depicted from Table 5.7.

Table 5.7 Performance of combine model

Forecasting **at Mundali** with KM-7 and BM-3								
Performance measures	MLFF		RBF		TS		LTF	
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation
$Q_{M(t)} = f(Q_{K(t-12)}, Q_{K(t-11)}, Q_{K(t-10)}, Q_{K(t-9)}, Q_{K(t-8)}, Q_{K(t-7)}, Q_{K(t-6)}, Q_{B(t-6)}, Q_{B(t-5)}, Q_{B(t-4)})$								
R^2	0.955	0.952	0.963	0.957	0.9779	0.976	0.938	0.905
RMSE	1337.5	1363.3	1303.8	1357.3	1270	1283.1	1655.5	1891.5
Efficiency	0.9348	0.9319	0.9381	0.9351	0.9441	0.9436	0.905	0.8806

Combination of best model of Khairmal (KM-7) and Barmul (BM-3) provides a best combination model (Case-III) for forecasting the flood at Mundali. The model performance statistics of this combined model is presented in Table 5.7. This combined model has been developed using LTF, MLP, RBF and TS fuzzy inference system. The LTF model efficiency is obtained as 0.905 during calibration and 0.8806 during validation (Table 5.7). The root mean square value is obtained as 1655.5 m^3/s during calibration and 1891.50 m^3/s during validation. Similarly, the coefficient of correlation is observed as 0.938 during calibration and 0.905 during validation.

In case of combined MLP model, the NS efficiency is obtained as 0.9348 during calibration and 0.9319 during validation. The root mean square value is obtained as 1337.5 m^3/s during calibration and 1363.30 m^3/s during validation. Similarly, the coefficient of correlation is observed as 0.955 during calibration and 0.952 during validation. The RBF model efficiency is obtained as 0.9381 during calibration and 0.9351 during validation. The root mean square value is obtained as 1303.8 m^3/s during calibration and 1357.30 m^3/s during validation. Similarly, the coefficient of correlation is observed as 0.963 during calibration and 0.957 during validation. Similarly, the TS model efficiency is obtained as 0.9441 during calibration and 0.9436 during validation. The root mean square value is obtained as 1270.0 m^3/s during calibration and 1283.10 m^3/s during validation. Similarly, the coefficient of correlation is observed as 0.9779 during calibration and 0.976 during validation.

In Case-III, the performance of MLP, RBF and TS is better than LTF. However the performance of LTF is also satisfactory. Also the RMSE values vary within 3 to 4% of the maximum peak value of 2008.

5.7.2.6 Model parsimony

The adequacy of inputs is tested by calculating the AIC and BIC statistics of the used 14 model combinations. The calculations are made as per the eq^n. 5.18 and 5.19, and recorded in Table 5.8. Generally, models having their AIC within 1 to 2 times of the minimum have substantial support and should receive consideration in making inferences.

AIC value within 4 to 7 times of the minimum have considerably less support, while models with their AIC >10 times above the minimum have either essentially no support and might be omitted from further consideration or at least fail to explain some substantial structure variation in the data. It is revealed from the analysis that the KM-7 model which has given a better performance also has given very less AIC and BIC values in comparison to other models with only Khairmal flow data and similarly BM-3 model has minimum AIC and BIC value under its category. In all cases TS model provides the lowest AIC and BIC values.

Table 5.8 AIC and BIC values (normalised) of all models

Cases		AIC			BIC		
	Model	MLFF	RBF	TS	MLFF	RBF	TS
Only Khairmal discharge	KM-1	1.041	1.038	1.027	1.041	1.038	1.027
	KM-2	1.037	1.033	1.023	1.038	1.033	1.024
	KM-3	1.035	1.030	1.022	1.035	1.031	1.023
	KM-4	1.029	1.028	1.017	1.030	1.028	1.017
	KM-5	1.027	1.025	1.016	1.028	1.025	1.017
	KM-6	1.025	1.022	1.013	1.026	1.023	1.014
	KM-7	1.018	1.017	1.01	1.021	1.020	1.012
	KM-8	1.022	1.021	1.012	1.024	1.023	1.014
	KM-9	1.022	1.021	1.012	1.024	1.023	1.014
Only Barmul discharge	BM-1	1.027	1.026	1.015	1.028	1.026	1.015
	BM-2	1.024	1.022	1.012	1.024	1.022	1.012
	BM-3	1.016	1.015	1.008	1.021	1.019	1.009
	BM-4	1.020	1.019	1.009	1.021	1.020	1.010
	BM-5	1.020	1.019	1.009	1.021	1.020	1.010
Khairmal and Barmul discharge	KBM	1.008	1.008	1.000	1.010	1.009	1.001

5.7.2.7 Uncertainty in model output

The above said models are further checked for predictive uncertainty using 'noise to signal ratio' statistics. The statistics are presented in calibration and validation range in Fig. 5.18. It is revealed that in all cases the statistics are well within 1. As per LTF analysis the difference between calibration and validation is wider, which shows the data used are noisy and LTF is unable to handle it very perfectly. The same data bears a good performance in TS. The differences in all KM models are more than BM models. The KBM combination shows a very less difference between calibration and validation. The KM-7, BM-3 and KBM combinations under TS analysis shows very less difference between calibration and validation. Thus these model combinations show TS models can handle the noisy data better than MLP and RBF models.

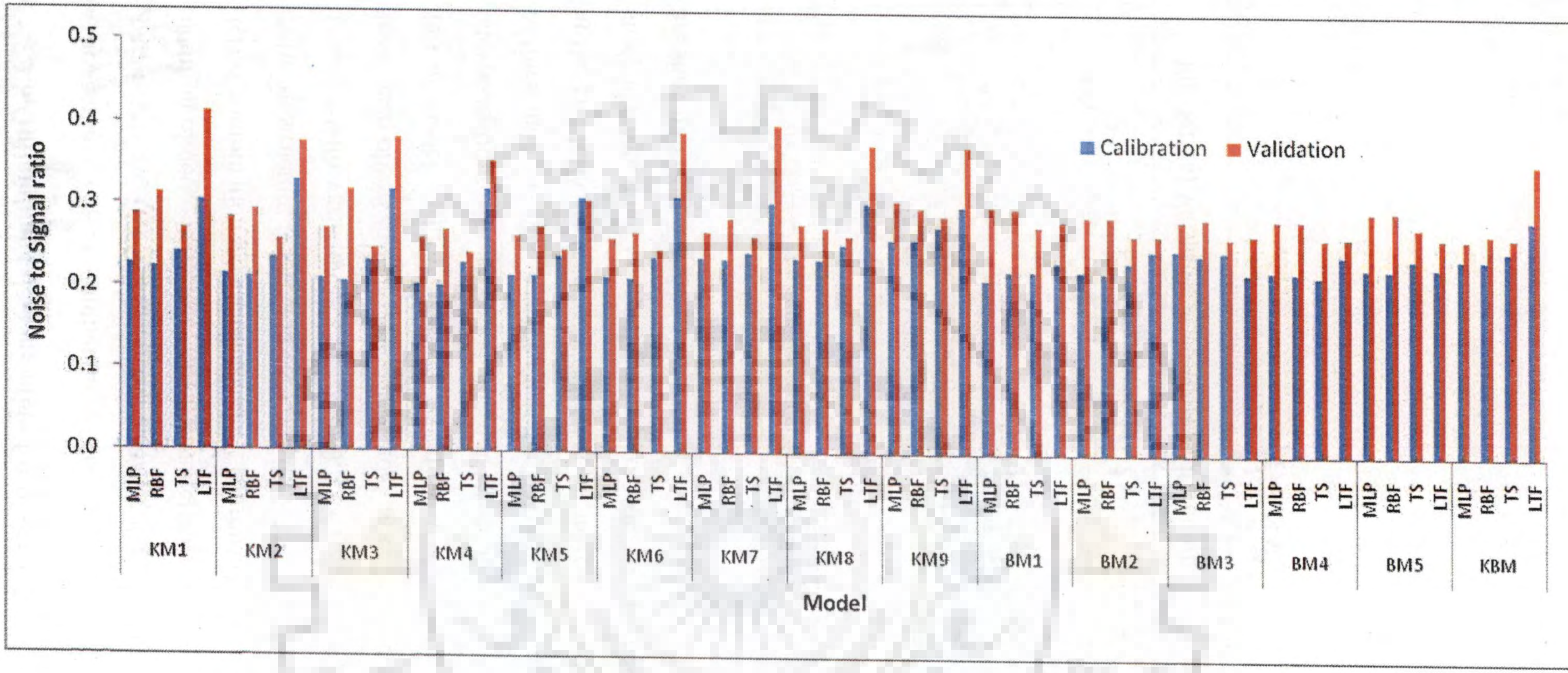


Fig. 5.18 Noise to signal ratio values of all 15 model combination during calibration and validation

5.7.2.8 Optimal model parameters for best model

The applied model combinations deliver their best performance after the optimal model parameters are fixed. As discussed earlier the performances of the ANN models are tested for addition of neurons. Besides other parameters, neurons are varied in this study from 1 to 12. The KBM combination gets stabilized at 7 neurons, KM-7 at 5 and BM-3 at 3 neurons. The optimal network structure for KBM is 10-7-1, for KM-7 is 7-5-1 and for BM-3 is 5-3-1. The RBF model is tested with variation of goal from 0.5 to 1.2 and spread from 1 to 2.2. Very minute changes have been observed with changes in goal whereas by changing spread significant variations are felt in model performances. So variation of Efficiency and RMSE with respect to spread is shown in figure. In TS model the cluster radius is kept varied from 0.2 to 0.8 to achieve the optimal model radius. It is seen that the KM-7 attains maximum performance at radius 0.4, BM-3 at 0.5 and KBM at 0.35, and after that no increase in performance occurs and the plot remains straight throughout. Optimal parameters used for the best models like KM-7, BM-3 and KBM are shown in (Fig. 5.19(a) to 5.21(b)).

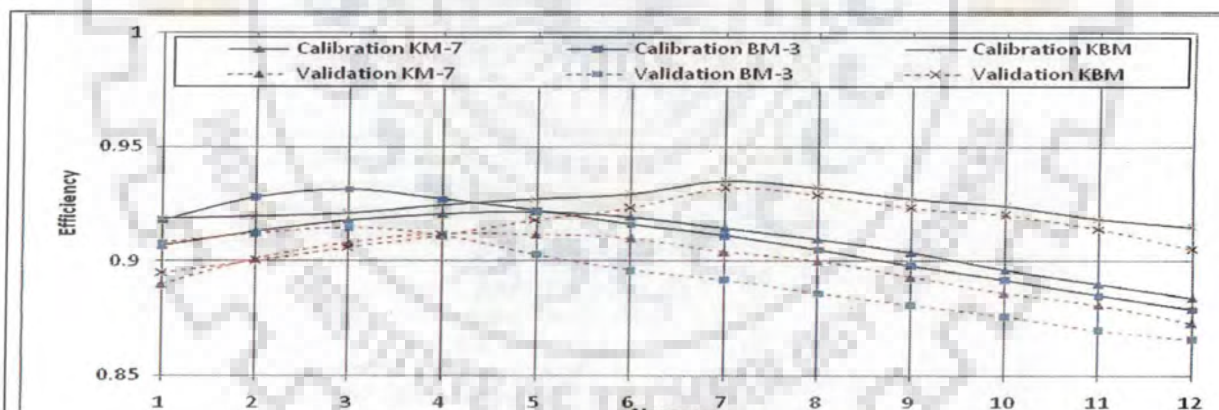


Fig. 5.19(a) Variation of efficiency with number of hidden neurons in 3 best combinations

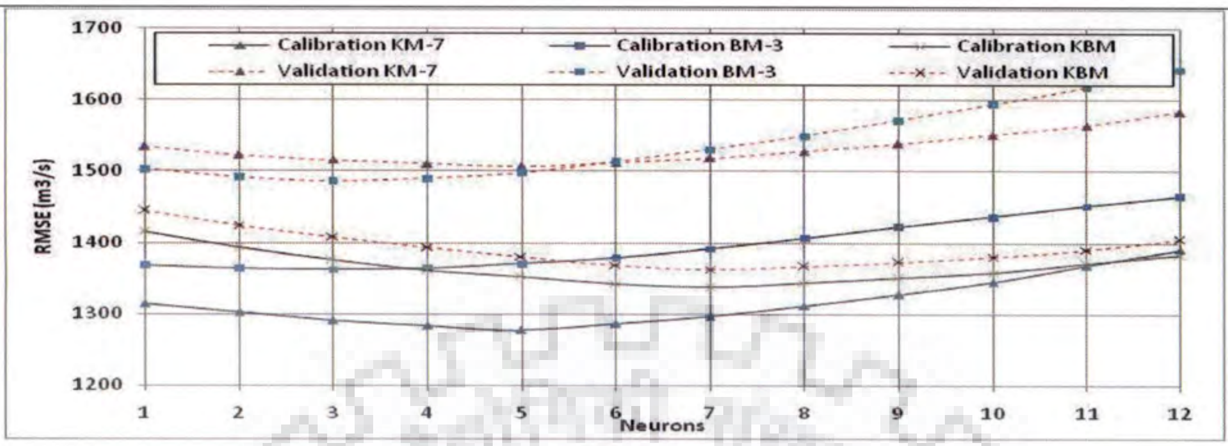


Fig. 5.19(b) Variation of RMSE with number of hidden neurons in 3 best combinations

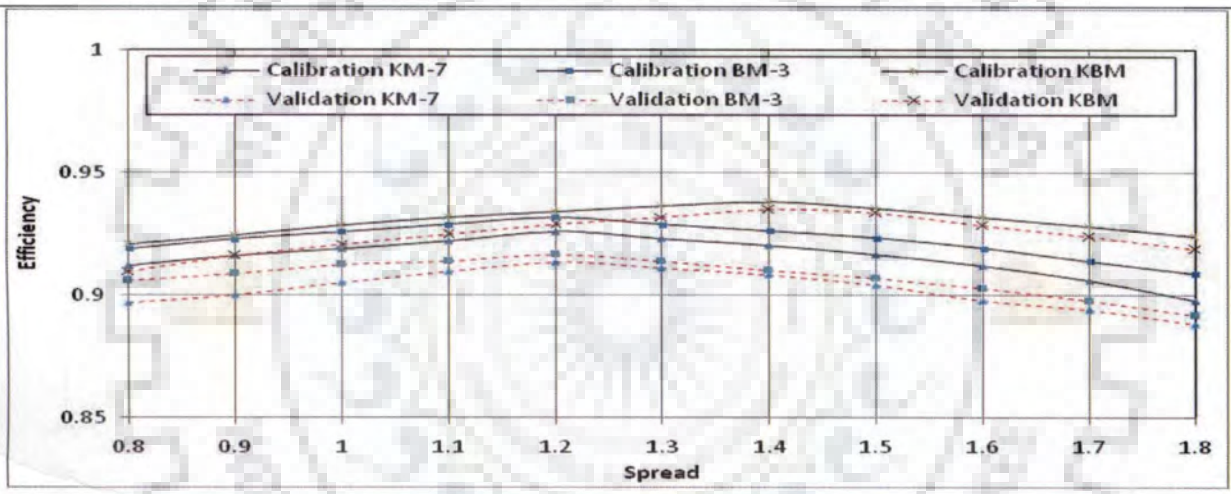


Fig. 5.20(a) Variation of efficiency with different spread in 3 best combinations

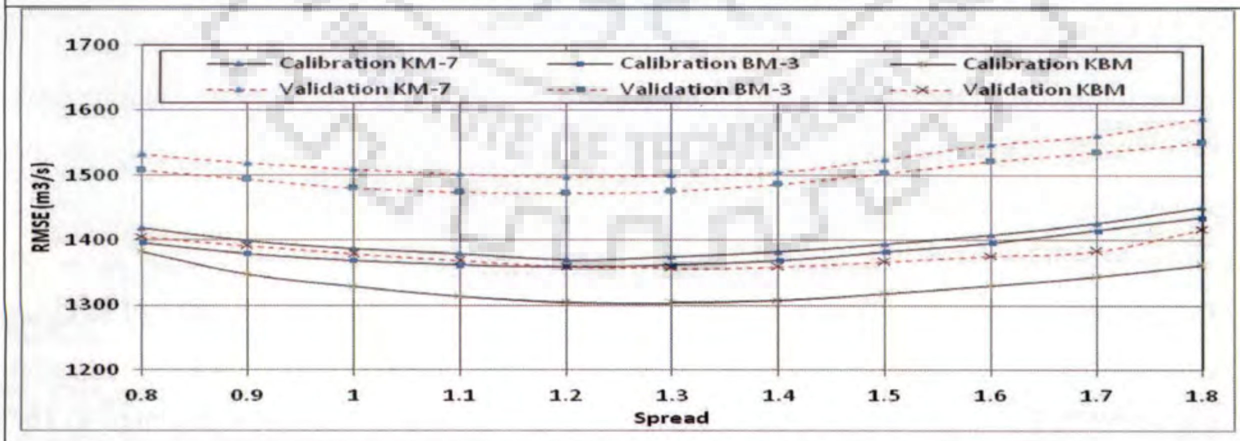


Fig. 5.20(b) Variation of RMSE with different spread in 3 best combinations

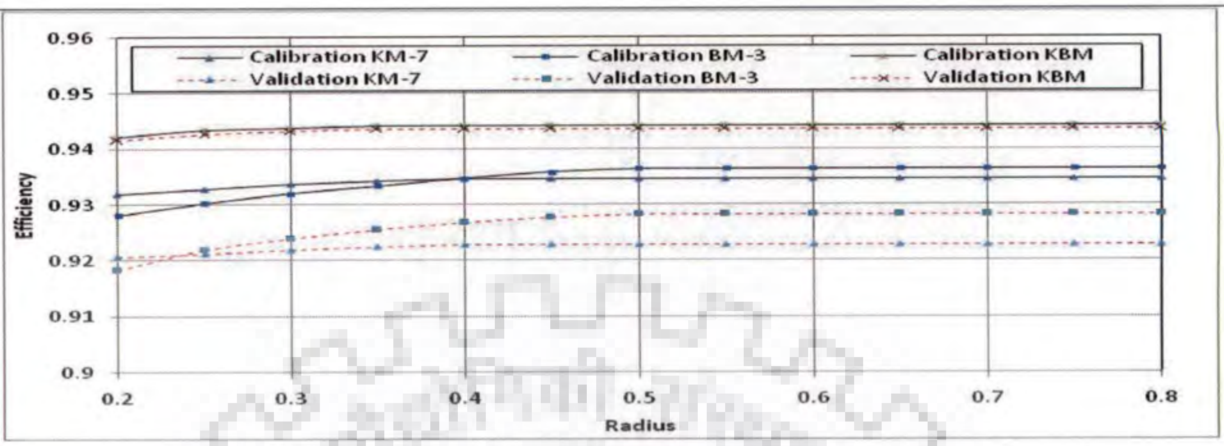


Fig. 5.21(a) Variation of Efficiency with different radius in 3 best combinations

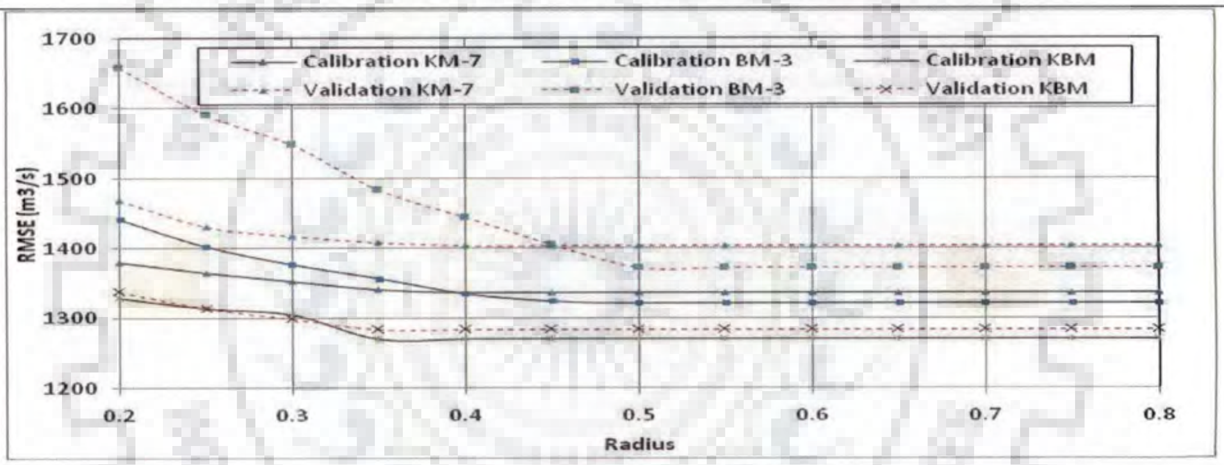


Fig. 5.21(b) Variation of RMSE with different radius in 3 best combinations

5.7.2.9 Comparison of peak floods forecasted by different models

The performances of these models are further checked with respect to their behavior during peak floods as well as total volume accumulated. For this 22 peak flood values were selected from the available 3 hourly discharge data from year 1996 to 2001 of Mundali G&D site. The forecasted discharges from 3 selected models (KM-7, BM-3 and KBM) are compared with the observed discharge and presented in Table 5.9, 5.10 and 5.11. It is seen that the performance of KBM is better than KM-7 and BM-3. As KM-7 is based on the information of 200 km. away and comparatively less catchment information is incorporated in it, so the results are bit inferior. The BM-3 model performances better than the KM-7 in peak flood prediction as in the case of BM-3 the forecasting is made basing on further downstream information. In KBM all information of upstream and intermediate stations are considered, so the result remains better than other combinations. In all cases, TS models show better result than ANN, RBF and LTF models. The results of these models are also compared with the time lag method which is presently used by DOWR, Orissa for flood forecasting. As the travel time of peaks considered by DOWR is 12-16 hours between Khairmal to Barmul and Barmul to Mundali all the predicted values are compared with the 12 hour and 15 hour peaks (as per DOWR norms) for models KM-7 and BM-3 and 24 and 30 hour peaks for model KBM. Performance of the model KM-7 indicates that the forecasted peak difference (in %) varies from -16.7 to 6.6, -11.0 to 5.3, -12.5 to 6.7, -43.2 to 15.6, -40.3 to 11.6 and -48.7 to 18.8 for MLP, RBF, TS, LTF, DOWR-24 and DOWR-30, respectively.

Performance of the model BM-3 indicates that the forecasted peak difference (in %) varies from 10.4 to -7.0, 9.2 to -8.6, 11.3 to -5.4, 12.4 to -43.1, 17.3 to -20.3 and 16.8 to -22.0 for MLP, RBF, TS, LTF, DOWR-12 and DOWR-15, respectively. Similarly, performance of the model KBM model indicates that the forecasted peak difference (in %) varies from 9.9 to 7.6, 9.5 to -8.6, 5.0 to -5.6, 9 to -66.5, 25.8 to -53.4,16.8 to -27.6,11.6 to -40.3 and 18.8 to -48.7 for MLP, RBF, TS, LTF, DOWR-12, DOWR-15, DOWR-24 and DOWR-30, respectively.

DOWR method (in BM-3 DOWR-12) is able to model only one peak of 2001 with less than 1% of difference between observed and computed discharge (flood peak of 2001 are next to 2008). However, other DOWR models are not able to forecast the 2001 peak flood more accurately. The 2001 peaks are comparatively better modeled by TS method in all cases. The 2008 peaks are also better modeled by TS method (in all) than DOWR method (in all). ANN and RBF methods also show much better result than DOWR method in all the three cases i.e. KM-7, BM-3 and KBM. However, the LTF method shows a comparatively very inferior result. Further, the TS models found very capable of modeling high or medium peaks rather than low peaks. The commonly used criterion for evaluating river flow forecasts in India considers ±20% variation between predicted and observed discharge values as reasonably accurate (CWC, 1989)

Table 5.9 Year wise peak differences between observed and modeled outputs for model KM-7

Year	Observed discharge (m^3/s)	Computed discharge (m^3/s)						Forecasted peak difference in %					
						DOWR(From Khairmal)						DOWR(From Khairmal)	
		MLP	RBF	TS	LTF	24 hr.	30 hr.	MLP	RBF	TS	LTF	24 hr.	30 hr.
1996	12514.4	12966	12252	12622	12239.3	12678.4	12678.4	3.6	-2.1	0.9	-2.2	1.3	1.3
1997	24279	24978	24598	24969	20452.3	18961	13867	2.9	1.3	2.8	-15.8	-21.9	-42.9
1998	23229.9	24590	24461	24273	20582.9	21225	22357	5.9	5.3	4.5	-11.4	-8.6	-3.8
1999	18043.7	18827	18587	18580	19520.8	16293.4	17376.2	4.3	3.0	3.0	8.2	-9.7	-3.7
2000	4717.4	3931.1	4222	4127	3809.9	5263.8	5603.4	-16.7	-10.5	-12.5	-19.2	11.6	18.8
2001	39154	37906	36493	38582	36051.4	35941	34809	-3.2	-6.8	-1.5	-7.9	-8.2	-11.1
	39869	37214	38543	38231	34005.9	36119.3	34809	-6.7	-3.3	-4.1	-14.7	-9.4	-12.7
	39630.6	37320	37827	38504	32487.5	32658.2	35941	-5.8	-4.6	-2.8	-18.0	-17.6	-9.3
2002	16229.3	17305	17024	16985	18760.8	14036.8	14716	6.6	4.9	4.7	15.6	-13.5	-9.3
2003	38200.6	37395	36542	38276	35285.9	34809	35714.6	-2.1	-4.3	0.2	-7.6	-8.9	-6.5
2004	21681.7	22818	22774	22184	18804.9	19866.6	19187.4	5.2	5.0	2.3	-13.3	-8.4	-11.5
2005	25362.4	23632	24082	25067	19461.5	19866.6	19187.4	-6.8	-5.4	3 .2	-23.5	-28.2	-31.1
2006	36318.5	33312	33912	36610	24296.4	32912.9	31130	-8.3	-6.6	0.8	-33.1	-9.4	-14.3
2007	20683.8	18070	18411	18725	14919.9	12338.8	10612.5	-12.6	-11.0	-9.5	-27.9	-40.3	-48.7
2008	44742.3	41352	41418	47754	34330.21	41035	39620	-7.6	-7.4	6.7	-23.3	-8.3	-11.4
	44742.3	41826	41466	46870	37975.96	41035	39620	-6.5	-7.3	4.8	-15.1	-8.3	-11.4
	44742.3	41795	40553	46101	43117.95	40695.4	41035	-6.6	-9.4	3.0	-3.6	-9.0	-8.3
	44487.6	40977	40500	43945	45196.53	40695.4	41035	-7.9	-9.7	-1.3	1.6	-8.4	-8.5
	44232.9	40770	40217	43081	46172.37	39337.0	40695.4	-7.8	-9.1	-2.6	4.4	-11.1	-8.0
	44006.5	40216	40222	42853	47702.72	37978.6	40695.4	-8.6	-8.6	-2.6	8.4	-13.7	-7.5
2009	24253.1	22017	22273	23232	16274.7	20998.6	17829	-9.2	-8.2	-4.2	-32.9	-13.4	-26.5
2010	19527.0	17414	17975	18245	11092.3	14716.0	13867	-10.8	-7.9	-6.6	-43.2	-24.6	-29.0

Table 5.10 Year wise peak differences between observed and modeled outputs for model BM-3

Year	Observed discharge (m^3/s)	Computed discharge (m^3/s)						Forecasted peak difference in %					
						DOWR (From Barmul)						DOWR (From Barmul)	
		MLP	RBF	TS	LTF	12 hr.	15 hr.	MLP	RBF	TS	LTF	12 hr.	15 hr.
1996	12514.4	13817	13660	12275	13326.32	12819.9	13980.2	10.4	9.2	-1.9	6.5	2.4	11.7
1997	24279	25995	25163	24777	27300.1	28469.8	28356.6	7.1	3.6	2.1	12.4	17.3	16.8
1998	23229.9	25146	24305	24082	23306.3	20683.6	18734.6	8.2	4.6	3.7	0.3	-11.0	-19.4
1999	18043.7	19275	19095	19089	18769.7	18456.8	18678	6.8	5.8	5.8	4.0	2.3	3.5
2000	4717.4	4982.8	5001.2	5249.5	5138.96	5306.3	5377	5.6	6.0	11.3	8.9	12.5	14.0
2001	39154	38266	38490	38795	42138.38	41671.8	42223.6	-2.3	-1.7	-0.9	7.6	6.4	7.8
	39869	41858	39013	40154	40986.6	40242.6	41671.8	5.0	-2.1	0.7	2.8	0.9	4.5
	39630.6	40849	38453	40802	40652.95	38884.2	40242.6	3.1	-3.0	3.0	2.6	-1.9	1.5
2002	16229.3	15096	14930	15427	15170.4	14984.4	15175.9	-7.0	-8.6	-5.4	-6.9	-8.2	-7.0
2003	38200.6	36483	36960	37141	35641.5	34490.6	34490.6	-4.5	-3.2	-2.8	-6.7	-9.7	-9.7
2004	21681.7	20912	20635	21001	17142.6	21324.1	21401.9	-3.5	-4.8	-3.1	-20.9	-1.6	-1.3
2005	25362.4	23817	24276	24517	16448	23559.8	23347.5	-6.1	-4.3	-3.3	-35.1	-7.1	-7.9
2006	36318.5	34519	34678	35440	27831.5	34313.8	33252.5	-5.0	-4.5	-2.4	-23.4	-5.5	-8.4
2007	20683.8	20065.9	19934.5	20967	14963	17036.6	17036.6	-3.0	-3.6	1.4	-27.7	-17.6	-17.6
2008	44742.3	42352	42754	43818	34330.21	36960	35960	-5.3	-4.4	-2.1	-23.3	-17.4	-19.6
	44742.3	42326	42870	43366	31975.96	39960.0	36960.0	-5.4	-4.2	-3.1	-28.5	-10.7	-17.4
	44742.3	42995	43301	45853	33117.95	38960	39960	-3.9	-3.2	2.5	-26.0	-12.9	-10.7
	44487.6	42277	42945	43200	35196.53	40960	38960	-5.0	-3.5	-2.9	-20.9	-7.9	-12.4
	44232.9	41970	42081	43117	36172.37	41960	40960	-5.1	-4.9	-2.5	-18.2	-5.1	-7.4
	44006.5	41116	42153	43022	37702.72	38592.1	41960	-6.6	-4.2	-2.2	-14.3	-12.3	-4.7
2009	24253.1	22910	23699	24627	13796.3	23687.1	23036.2	-5.5	-2.3	1.5	-43.1	-2.3	-5.0
2010	19527.0	21067	20871	20603	12655.2	15557.9	15239.6	7.9	6.9	5.5	-35.2	-20.3	-22.0

Table 5.11 Year wise peak differences between observed and modeled outputs for model KBM

Year	Observed discharge (m3/s)	Computed discharge (m3/s)								Forecasted peak difference in %							
						DOWR (From Barmul)		DOWR (From Khairmal)						DOWR (From Barmul)		DOWR (From Khairmal)	
		MLP	RBF	TS	LTF	12 hr.	15 hr.	24hr.	30hr.	MLP	RBF	TS	LTF	12 hr.	15 hr.	24hr.	30hr.
1996	12514.4	12994	13019	12578	13637	12819.9	13980.2	12678	12678.4	3.8	4.0	0.5	9.0	2.4	11.7	1.3	1.3
1997	24279	22869	23292	24424	15000	28469.8	28356.6	18961	13867	-5.8	-4.1	0.6	-38.2	17.3	16.8	-21.9	-42.9
1998	23229.9	21942	21875	23243	19159	20683.6	18734.6	21225	22357	-5.5	-5.8	0.1	-17.5	-11.0	-19.4	-8.6	-3.8
1999	18043.7	18785	18664	18480	19115	18456.8	18678	16293	17376.2	4.1	3.4	2.4	5.9	2.3	3.5	-9.7	-3.7
2000	4717.4	5182.5	5165	4951.1	4524	5306.3	5377	5263.8	5603.4	9.9	9.5	5.0	-4.1	12.5	14.0	11.6	18.8
2001	39154	40506	40456	40132	31413	41671.8	42223.6	35941	34809	3.5	3.3	2.5	-19.8	6.4	7.8	-8.2	-11.1
	39869	39305	39397	40048	31203	40242.6	41671.8	36119	34809	-1.4	-1.2	0.4	-21.7	0.9	4.5	-9.4	-12.7
	39630.6	38251	38687	39382	30899	38884.2	40242.6	32658	35941	-3.5	1.1	1.8	-21.5	25.8	3.5	-18.8	10.1
2002	16229.3	14994	14942	15388	14306	14984.4	15175.9	14037	14716	-7.6	-8.6	-5.6	-12.5	-8.7	-7.0	-14.4	-10.8
2003	38200.6	36580	37002	39699	28650	34490.6	34490.6	34809	35714.6	-4.2	-3.1	3.9	-25.0	-9.7	-9.7	-8.9	-6.5
2004	21681.7	21405	21370	21853	16131	21324.1	21401.9	19867	19187.4	-1.3	-1.4	0.8	-25.6	-1.6	-1.3	-8.4	-11.5
2005	25362.4	24161	24421	24875	18416	23559.8	23347.5	19867	19187.4	-4.7	-3.7	-1.9	-27.4	-7.1	-7.9	-21.7	-24.3
2006	36318.5	34813	35691	36815	16809	34313.8	33252.5	32913	31130	-4.1	-1.7	1.4	-53.7	-5.5	-8.4	-9.4	-14.3
2007	20683.8	20294	20187	20901	18163	17036.6	17036.6	12338.8	10612.5	-1.9	-2.4	1.1	-12.2	-17.6	-17.6	-40.3	-48.7
2008	44742.3	41952	43754	45418	34330	36960	35960	41035	39620	-6.2	-2.2	1.5	-23.3	-17.4	-19.6	-8.3	-11.4
	44742.3	42026	43870	45466	37976	39960.0	36960.0	41035	39620	-6.1	-1.9	1.6	-15.1	-10.7	-17.4	-8.3	-11.4
	44742.3	41795	43301	45553	33118	38960	39960	40695.4	41035	-6.6	-3.2	1.8	-26.0	-12.9	-10.7	-9.0	-8.3
	44487.6	42007	42945	45500	35197	40960	38960	40695.4	41035	-5.6	-3.7	2.4	-20.4	-10.0	-13.5	-9.7	-8.5
	44232.9	41770	42081	45217	36172	41960	40960	39337.0	40695.4	-5.6	-4.9	2.2	-18.2	-5.1	-7.4	-11.1	-8.0
	44006.5	41816	42853	44222	37703	38592.1	41960	37978.6	40695.4	-5.0	-2.8	0.5	-14.3	-14.4	-5.3	-14.4	-8.7
2009	24253.1	23513	23929	23789	8442	23687.1	23036.2	20998.6	17829	-3.1	-1.4	-1.9	-66.5	-6.7	-5.1	-14.1	-30.6
2010	19527.0	18468	18423	19920	7433	15557.9	15239.6	14716.0	13867	-5.4	-6.0	2.1	-60.7	-53.4	-27.6	-31.6	-38.5

5.7.2.10 Comparison of 6-day flood hydrograph of 2008 forecasted by different models

Testing the performance of each model depends on how it behaves for predicting the peaks. To check instant behavior of peak hydrographs and response of models towards this 6 day observed and modeled hydrographs are drawn on a trial basis again for the highest recorded flood event of 2008. The highest recorded flood at Mundali is 44742 m^3/s on of 21st September 2008. So a hydrograph starting from 18th September to 23rd September is drawn for all KM-7, BM-3 and KBM models (Fig. 5.22 (a) to Fig. 5.22(c)). It is again revealed that the TS model of KBM combination perfectly forecasts the observed hydrograph at Mundali. The KBM-MLP and KBM-RBF are a little under predicted whereas LTF shows a poor performance. Simultaneously, the existence of KM-7 and BM-3 models also give very influential results and can also be referred in absence of KBM or as best alternatives.

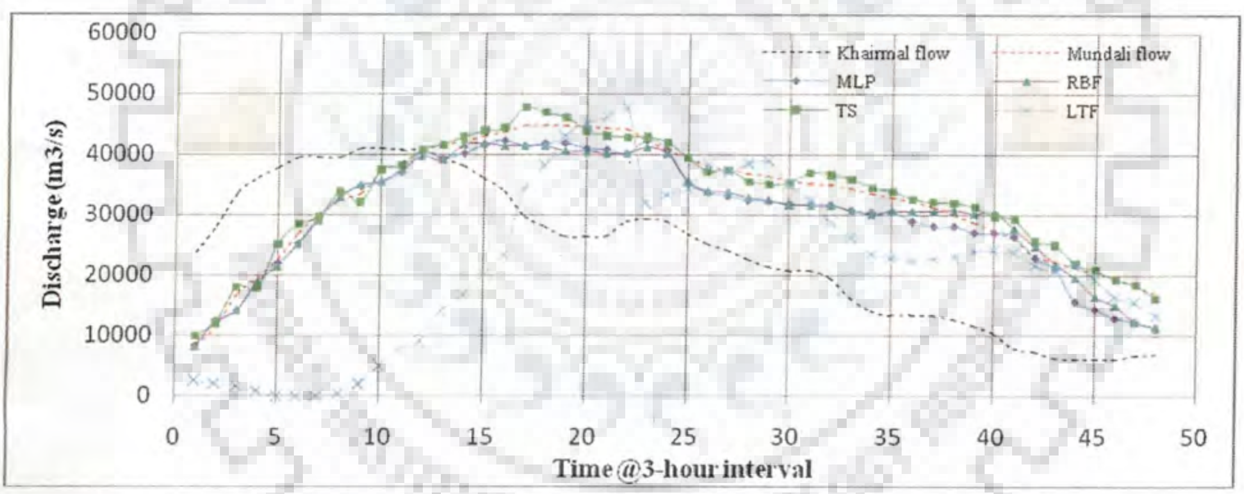


Fig. 5.22(a) Comparison between observed and computed 6-day flood hydrograph of 2008 at Mundali for KM-7 combination

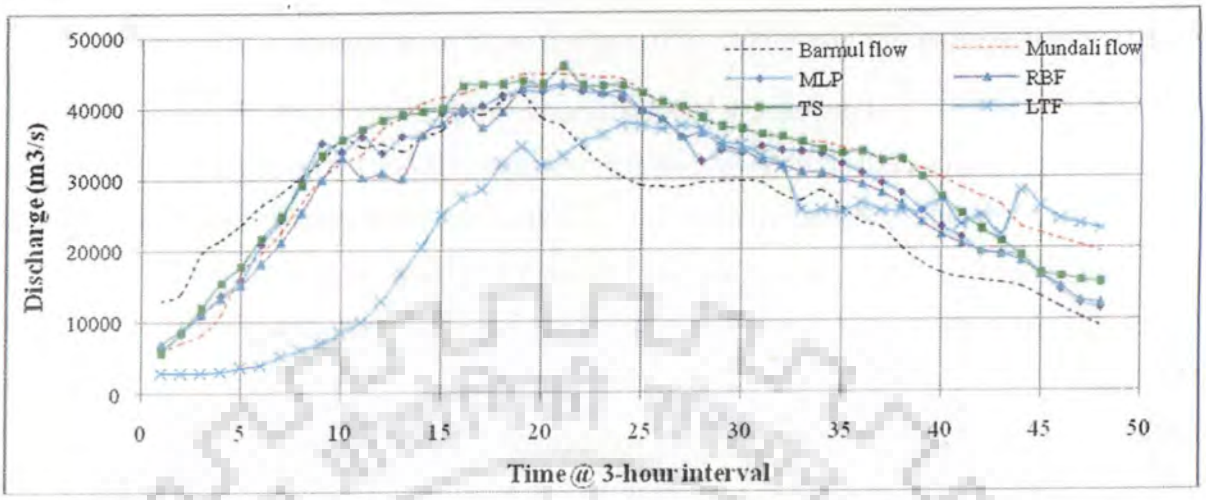


Fig. 5.22(b) Comparison between observed and computed 6-day flood hydrograph of 2008 at Mundali for BM-3 combination

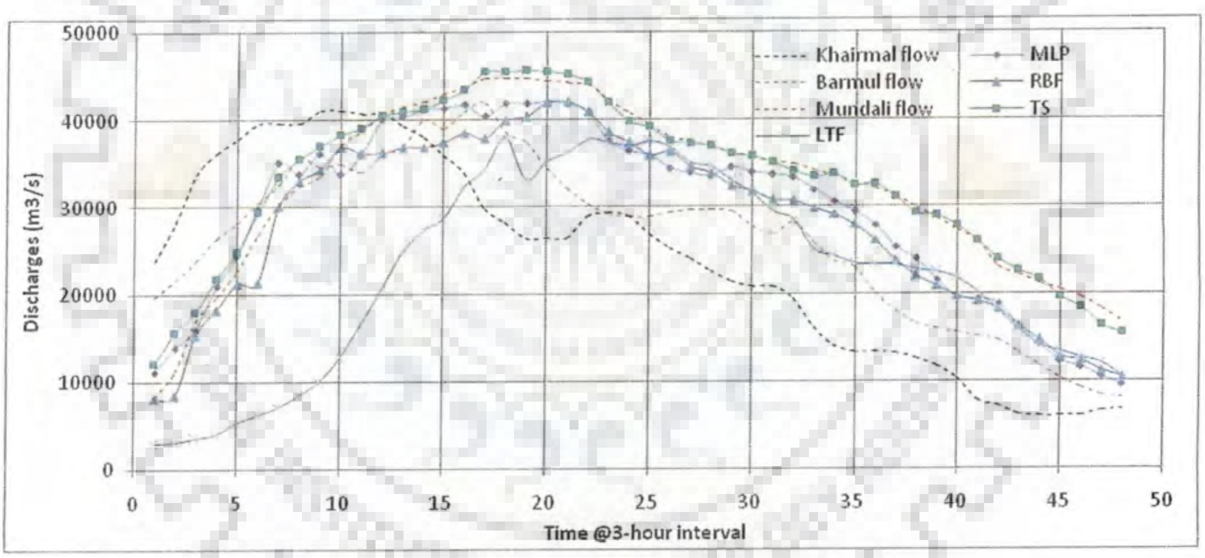


Fig. 5.22(c) Comparison between observed and computed 6-day flood hydrograph of 2008 at Mundali for KBM combination

5.7.2.11 Volumetric check of 2008 flood

Generally checking of runoff volumes are done on yearly basis for rainfall-runoff models. In this study, the volume of total yearly runoff is also checked for cross-checking the performance of each model. Instead of testing it for all years in this study it is checked for only year 2008 as this year has seen the heaviest flood in the history so far after that of 1982. The total runoff volume in million cubic meter (mcm) accumulated in the year 2008 is 57916.4. When this is modeled by KM-7, BM-3 and KBM, the accumulated volumes through TS approach has given a close answer upto 96% of observed value in KBM, 95.9% in BM-3 and 90.0% in KM-7 approach (Table 5.12). The RBF and MLP models go very close to each other. The result of KM-7 model is a bit inferior to BM-3 and KBM and also the result of LTF is poor in comparison to MLP, RBF and TS.

The accumulation of runoffs is plotted in Fig. 5.23 (a) to 5.23 (c) of all 3 models for the flood of 2008. It is revealed from the Table 5.12 that although the accumulated volumes are very close to observed volumes the plots in Fig. 5.23 (a) to 5.23 (c) showing a identifiable difference at the end but are very close when sudden rise occurs during high floods. The reason could be the peaks are getting modeled perfectly than medium and low peaks. In the TS approach peak modeling has been perfectly done in KBM model combination. The same is also supported from the result of Table 5.11 of KBM.

Table 5.12 Accumulated runoff volume for 3 best model combinations for the year 2008

Model	Observed volume (mcm)	MLP		RBF		TS		LTF	
		Total vol. (mcm)	% of total vol.	Total vol. (mcm)	% of total vol.	Total vol. (mcm)	% of total vol.	Total vol. (mcm)	% of total vol.
KM-7	57916.4	51791.3	89.4	51454.7	89.0	52110.3	90.0	46504.9	80.0
BM-3		53615.2	92.6	53698.5	93.0	55486.6	95.9	51216.5	88.0
KBM		53979.9	93.2	54195.3	94.0	55548.1	96.0	52652.9	91.0

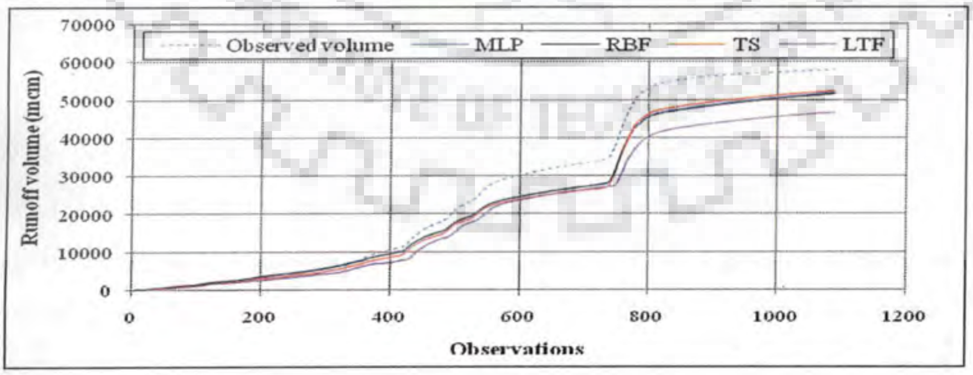


Fig. 5.23(a) Runoff volume using KM-7 model

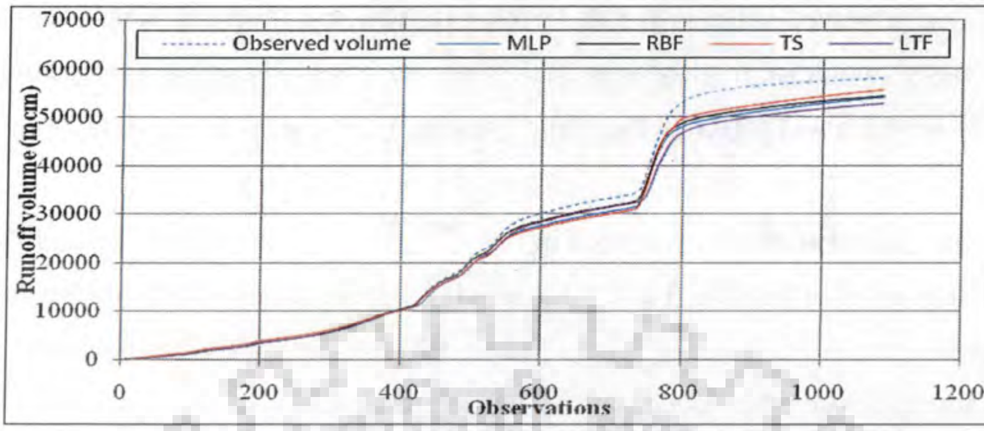


Fig. 5.23(b) Runoff volume using BM-3 model

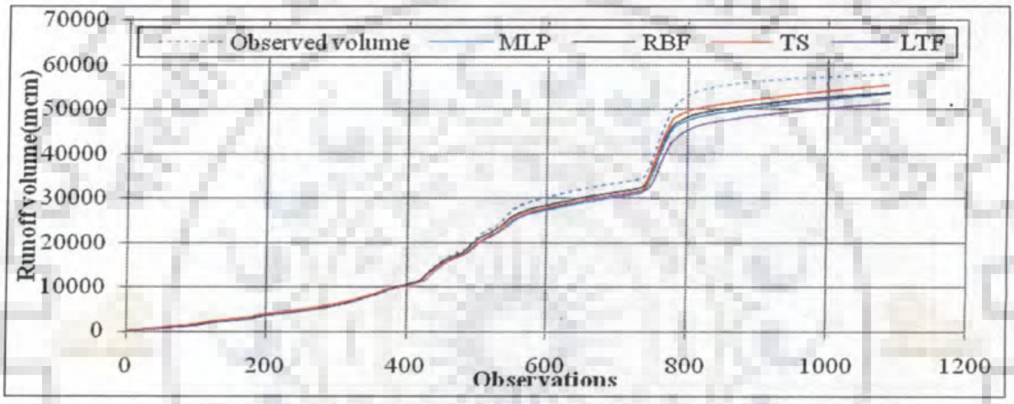


Fig. 5.23(c) Runoff volume using KBM model

So considering all the factors and result analysis the KBM model combination with TS approach is considered as best among all. However, the BM-3 and KM-7 are next alternatives that will be workable as per the availability of the information.

5.7.2.12 Performance with lead times

Thus KBM model has been recommended for flood forecasting in this study as it gives very good performance. But simultaneously KM-7 has also the advantage of predicting a bit earlier so can be carried out without waiting for Barmul information or when it is not available. BM-3 also can be favoured when Khairmal information is not available. Thus, all 3 models are selected and will be applied as per the requirement and availability of information. So the performance of all 3 models under different lead times are tested through MLP, RBF and TS approach and recoded in Table 5.13(a) to 5.13(c). The forecasting has been provided up to a 48 hour lead time. No further lead times are considered as efficiency and other performances fall drastically beyond 48 hour lead time.

For TS method in KBM the forecasting efficiency varies from 94.36 % to 86.7% for 21 to 48 hour in advance during validation. For the same time frame MLP and RBF results vary from 93.51% to 85.39% and 93.39% to 84.17%.

Table 5.13(a) Performance of KM-7 model under different lead times

Lead Time (hours)	Performance	MLP		RBF		TS	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
21	R	0.9626	0.9435	0.9628	0.9493	0.9768	0.9644
	RMSE	1377.4	1507.1	1370.1	1497	1336.4	1403.4
	Eff.	0.9215	0.9114	0.926	0.9136	0.9346	0.9228
24	R	0.9563	0.9371	0.9607	0.9436	0.9755	0.9623
	RMSE	1428.5	1568.7	1420.2	1529.9	1366.8	1435.8
	Eff.	0.9172	0.9064	0.9217	0.9093	0.9307	0.9201
27	R	0.9489	0.9298	0.9587	0.9413	0.9728	0.9555
	RMSE	1483.1	1612	1454.3	1567.6	1395.6	1500.5
	Eff.	0.9122	0.9007	0.9168	0.9043	0.9271	0.9139
30	R	0.9426	0.9243	0.956	0.9383	0.9701	0.945
	RMSE	1553.7	1696.9	1523.7	1644.3	1444.7	1584.7
	Eff.	0.9045	0.8926	0.9105	0.8979	0.9227	0.9071
33	R	0.9336	0.915	0.9526	0.9324	0.9677	0.94
	RMSE	1665.6	1834.5	1613.8	1800.6	1542.1	1738.5
	Eff.	0.8959	0.8816	0.9025	0.8877	0.9169	0.8972
36	R^2	0.9282	0.9018	0.9486	0.9279	0.9631	0.9333
	RMSE	1787.3	2002	1738.7	1941.3	1627.3	1889.7
	Eff.	0.8869	0.8687	0.8939	0.8738	0.9089	0.8855
39	R^2	0.9237	0.889	0.9435	0.9205	0.9589	0.9266
	RMSE	1963.7	2232.3	1883.2	2157.9	1776.8	2095
	Eff.	0.8757	0.8447	0.8828	0.8542	0.8987	0.8684
42	R	0.9173	0.8723	0.9377	0.9103	0.9516	0.9162
	RMSE	2181.5	2437.1	2070.6	2394.9	1981.1	2344.5
	Eff.	0.8497	0.8232	0.8667	0.8297	0.8867	0.8403
45	R	0.9025	0.8572	0.9242	0.8943	0.9397	0.902
	RMSE	2444.2	2631.3	2326	2605.1	2247.6	2573.5
	Eff.	0.8292	0.7912	0.8405	0.7974	0.8614	0.8075
48	R	0.8854	0.8419	0.9094	0.8709	0.9285	0.8767
	RMSE	2723.2	2894.3	2561.5	2823.8	2465.6	2782.5
	Eff.	0.8013	0.7537	0.822	0.7639	0.8473	0.7751

For TS method in BM-3 the forecasting efficiency varies from 92.83 % to 80.62% for 15 to 42 hour in advance during validation. For the same time frame MLP and RBF results vary from 91.45% to 77.57% and 91.67% to 78.69%.

Table 5.13(b) Performance of BM-3 model under different lead times

Lead Time (hours)	Performance	MLP		RBF		TS	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
15	R	0.9665	0.9606	0.9686	0.9609	0.9681	0.9655
	RMSE	1362.4	1486.5	1360.2	1473.5	1321.3	1371.6
	Eff.	0.9314	0.9145	0.932	0.9167	0.9364	0.9283
18	R	0.9631	0.9571	0.9657	0.9526	0.9665	0.9629
	RMSE	1412.5	1531.2	1393.2	1503.9	1347.8	1415.8
	Eff.	0.9262	0.9124	0.9287	0.9167	0.9327	0.9241
21	R	0.9589	0.9498	0.963	0.9483	0.9628	0.9592
	RMSE	1468.1	1592	1444.3	1551.6	1385.6	1460.5
	Eff.	0.9212	0.9087	0.9248	0.9127	0.9291	0.9218
24	R	0.9526	0.9408	0.961	0.9453	0.9601	0.955
	RMSE	1526.7	1666.9	1503.7	1614.3	1424.7	1524.7
	Eff.	0.9185	0.9052	0.9205	0.9099	0.9257	0.9149
27	R	0.9488	0.9335	0.957	0.9324	0.9577	0.951
	RMSE	1605.6	1754.5	1559.8	1695.6	1509.1	1608.5
	Eff.	0.9149	0.8966	0.9175	0.9017	0.9219	0.9082
30	R^2	0.9432	0.9248	0.953	0.9219	0.9531	0.9412
	RMSE	1744.3	1862	1633.7	1781.3	1587.3	1703.7
	Eff.	0.9069	0.8767	0.9109	0.8898	0.9139	0.8955
33	R	0.9382	0.9155	0.948	0.9165	0.9509	0.9346
	RMSE	1863.7	1992.3	1753.2	1907.9	1696.8	1825
	Eff.	0.8927	0.8497	0.9028	0.8662	0.9077	0.8764
36	R^2	0.9283	0.9023	0.941	0.9003	0.9456	0.9222
	RMSE	2051.5	2157.1	1860.6	2024.9	1781.1	1944.5
	Eff.	0.8787	0.8312	0.8867	0.8417	0.8967	0.8523
39	R	0.9156	0.8872	0.9282	0.8863	0.9347	0.9115
	RMSE	2241.2	2371.3	2016	2205.1	1917.6	2093.5
	Eff.	0.8562	0.8062	0.8655	0.8174	0.8773	0.8335
42	R	0.894	0.8603	0.9144	0.8609	0.9235	0.8857
	RMSE	2433.2	2644.8	2207.5	2423.8	2100.6	2282.5
	Eff.	0.8313	0.7757	0.844	0.7869	0.8584	0.8062

For TS method in KBM the forecasting efficiency varies from 94.36 % to 86.7% for 15 to 42 hour in advance during validation. For the same time frame MLP and RBF results vary from 93.51% to 85.39% and 93.39% to 84.17%.

Table 5.13(c) Performance of KBM model under different lead times

Lead Time (hours)	Performance	MLP		RBF		TS	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
15	R	0.955	0.952	0.963	0.957	0.9769	0.976
	RMSE	1337.5	1363.3	1303.8	1357.3	1270	1283.1
	Eff.	0.9348	0.9339	0.9381	0.9351	0.9441	0.9436
18	R	0.9539	0.9501	0.9617	0.9536	0.9745	0.9736
	RMSE	1372.5	1408.7	1343.2	1389.9	1306.8	1318.3
	Eff.	0.9302	0.9304	0.9347	0.9317	0.9417	0.9405
21	R	0.9529	0.9481	0.9587	0.9513	0.9718	0.9705
	RMSE	1418.1	1440	1384.3	1427.6	1315.6	1360.5
	Eff.	0.9292	0.9267	0.9318	0.9283	0.9381	0.9379
24	R	0.9536	0.9463	0.9561	0.9484	0.9661	0.964
	RMSE	1446.7	1486.9	1413.7	1464.3	1334.7	1384.7
	Eff.	0.9265	0.9206	0.9285	0.923	0.9347	0.9331
27	R	0.9508	0.9429	0.953	0.9414	0.9617	0.9581
	RMSE	1500.6	1550.5	1459.8	1535.6	1362.1	1417.6
	Eff.	0.924	0.9126	0.9258	0.9177	0.9299	0.9287
30	R^2	0.9462	0.9368	0.9456	0.9339	0.9531	0.9462
	RMSE	1564.3	1642	1508.7	1611.3	1397.3	1469.7
	Eff.	0.9199	0.9017	0.9219	0.911	0.9249	0.9245
33	R	0.9401	0.931	0.9385	0.9245	0.9409	0.9336
	RMSE	1603.7	1722.3	1543.2	1677.9	1436.8	1525
	Eff.	0.9137	0.8927	0.9158	0.9032	0.9187	0.9184
36	R	0.9323	0.9203	0.9277	0.9133	0.9306	0.9193
	RMSE	1675	1871.1	1610.6	1804.9	1501.1	1596.4
	Eff.	0.907	0.8752	0.9087	0.8927	0.9117	0.906
39	R	0.9178	0.9032	0.9142	0.8963	0.9187	0.9075
	RMSE	1755	2011.3	1696	1905.1	1547.6	1673.5
	Eff.	0.89	0.8612	0.898	0.8754	0.9034	0.8875
42	R	0.8962	0.8719	0.8984	0.8739	0.9005	0.8909
	RMSE	1863	2261.3	1797.5	2123.8	1595.6	1781.6
	Eff.	0.8663	0.8417	0.8736	0.8539	0.8893	0.867

The variations of RMSE and efficiency at different lead time for the 3 model combinations are provided in Fig. 5.24(a) and 5.25(c). It is evident from these figures that RMSE increases in lead time and efficiency decreases with it.

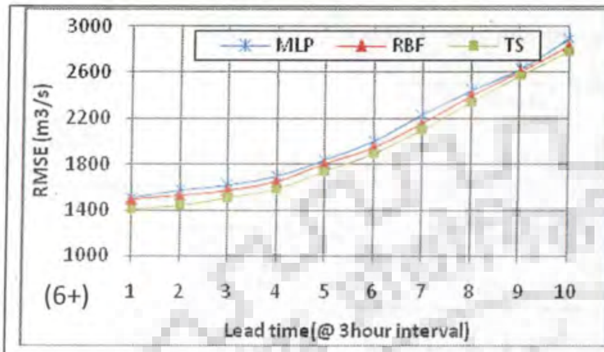


Fig. 5.24(a) Variation of RMSE at different lead times in KM-7 model during validation

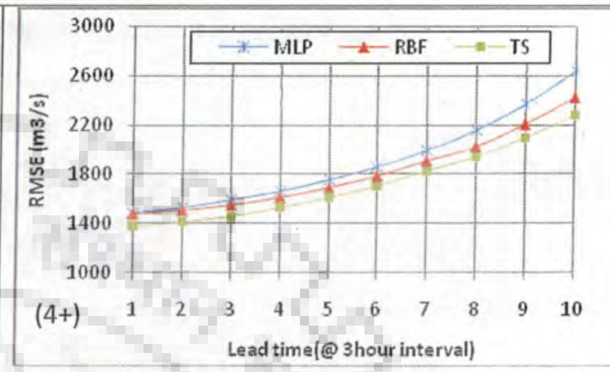


Fig. 5.24(b) Variation of RMSE at different lead times in BM-3 model during validation

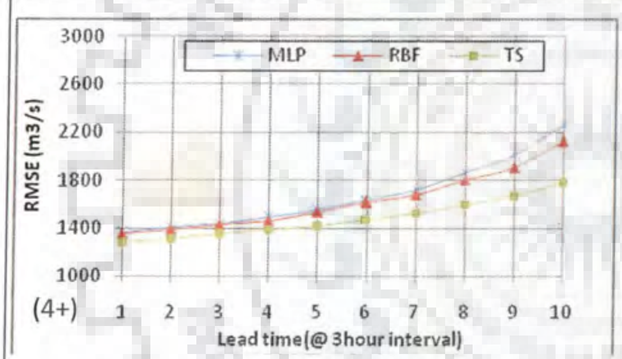


Fig. 5.24(c) Variation of RMSE at different lead times in KBM model during validation

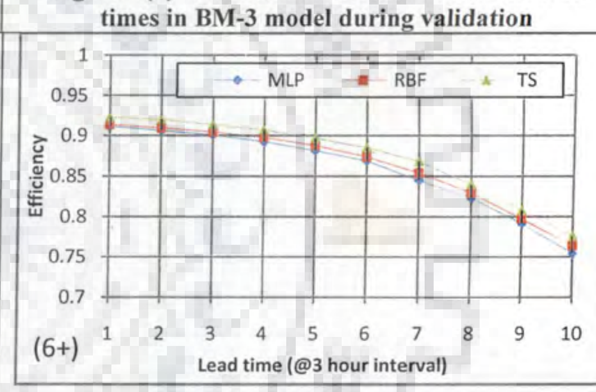


Fig. 5.25(a) Variation of Efficiency at different lead times in KM-7 model during validation

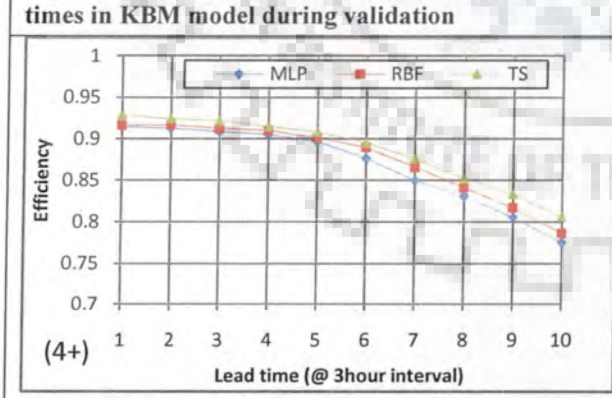


Fig. 5.25(b) Variation of Efficiency at different lead times in BM-3 model during validation

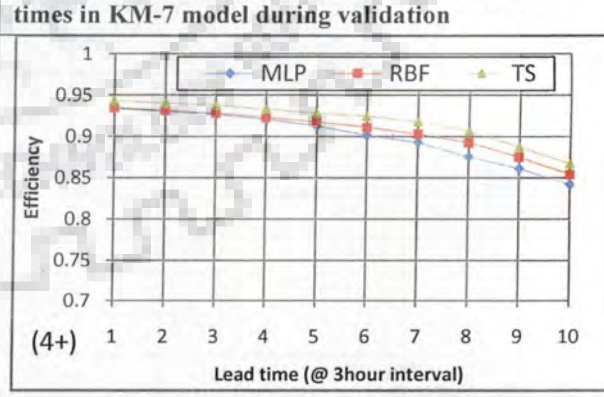


Fig. 5.25(c) Variation of Efficiency at different lead times in KBM model during validation

5.7.2.13 Discussion on application of soft computing models

A flood forecasting system basically depends on extensive and timely data collection, its transmission and application of inputs to a workable model and then finally dissemination of the output to authorities and to end user. In case of Mahanadi basin no doubt the data collection network and transmission procedures are improving in present days, IMD is also providing a lot of support for acquiring short interval data the short fall remains in development of a workable model. In that context existing model of DOWR which works on the daily rainfall information and Hirakud dam release has many disadvantages like

i) the existing method does not provide much accuracy

ii) it depends upon the rainfall information from different gauges

iii) does not provide lead time of forecast

In contrast to the above model, the developed soft computing models only require the discharge information at base station. Simultaneously it provides a good lead time.

5.8 CONCLUSIONS

In a flood prone basin like Mahanadi controlling floods through structural measures is a difficult and costly task. The non-structural measures like flood forecasting is a better solution towards sufferings of eight coastal districts of Orissa. The Hirakud dam controls only 83000 km^2 catchment and nearly 58000 km^2 downstream area remains uncontrolled. Existing flood forecasting system is based on time lag approach. Thus development of a flood forecasting model using soft computing technique is expected to improve the existing forecast quite significantly. In this study both peak discharge as well as 3-hourly discharges at Mundali have been forecasted. The following conclusions are drawn from this study:

i) Statistical, ANN and Fuzzy logic models are applied for forecasting of flows at Barmul and Mundali. The performance of TS-fuzzy method is better than MLFF-ANN, RBF-ANN and statistical methods.

ii) The studies indicate that RBF performs better than MLP in forecasting of flood flows in lower reach of Mahanadi basin.

iii) For forecasting of peak flows 3 models viz. Khairmal-Mundali (KM), Barmul-Mundali (BM) and Khairmal-Barmul-Mundali (KBM) have been developed. Performance wise KBM is better than BM and BM is better than KM. However, lead time is maximum in case of KM and performance of KM is also satisfactory.

iv) For forecasting of 3-hourly flow data, three recommended models are KM-7, BM-3 and KBM. Although, KBM is the best model out of all 15 model combinations, KM-7 and BM-3 can be used as alternate models.

v) The lead time is maximum case of KM-7 model. Hence, Model KM-7 is recommended for the situations where lead time of 24-32 is required, even though its performance is slightly inferior to BM-3 and KBM models.

5.9 SCOPE FOR FURTHER IMPROVEMENT

The study has a wide scope for further work in the area. The possible directions in which further work can be undertaken are listed below:

i) The flood forecasting is done in the downstream catchment using 3-hourly discharge data of 3 downstream stations of Hirakud reservoir. Non-availability of similar duration of rainfall, discharge of other tributaries and information of Hirakud release restricted the study to use the discharge data of 3 stations only. Further, flood forecasting studies may be taken up with the help of the forecasted rainfall to achieve a better lead time.

ii) A hydro-meteorological data observation network should be established in the Mahanadi basin in order to develop a flood forecasting model for the entire Mahanadi basin for the better management of floods.

iii) There is always a conflict of superiority between RBF-ANN and MLFF-ANN network. Although in our study RBF has given a better performance, it requires further extensive testing in different watersheds.

CHAPTER 6 - RAINGAUGE NETWORK DESIGN OF KANTAMAL SUBBASIN OF LOWER MAHANADI FOR FLOOD FORECASTING

6.1 BACKGROUND

Timely collection and transmission of the hydrological information from the gauging site to the flood forecasting station is a very important task for the successful operation of real time flood forecasting models. Sometimes it becomes difficult to collect information from all raingauges (RG) either due to instrumental disorder, difficulty in communication, inability to take readings or many other reasons during flood times. Therefore, it is important to find out key RG networks capable of forecasting the flood without compromising much with the forecasting accuracy. In this chapter a procedure for the design of the key rain gauge network particularly important for the flood forecasting is discussed and demonstrated through a case study.

6.2 INTRODUCTION

Information of rainfall is the primary requirement of all flood forecasting models. It is always not possible to gather information from all RGs. The reasons could be many. In particular, during flood time there may be chances of failure, breaking, non-recording of RGs, difficulty in transmission of information etc. In large catchments these uncertainties are more prominent.

The research for establishing key RG network is always in the hunt. Earlier Kagan (1966) had suggested a procedure to compute the error in estimation of aerial rainfall which could be used in estimation of key network density of RGs. Hall (1972) suggested a rational method for determination of key station network. Morin et al. (1979) advocated the use of principal component analysis in conjunction with optimal interpolation for RG network design. Sreedharan and James (1983) used the spatial correlation technique proposed by Kagan for design of RG network.

The real world problem is always complex but it requires decision. Saaty (1980) has introduced Analytic Hierarchy Process (AHP) for solving the complex decision oriented problems. It can make decisions involving much kind of concerns including planning, setting priorities, selecting the best among number of alternatives and allocating resources. An AHP can effectively deal with both qualitative and quantitative factors in multiple criteria decision environments. It is an important decision tool because of its ability to synthesize multi-attributed scenarios and provide diagnostic information, which enables decision makers to better understand the behavioral processes underlying choices. From its inception, and arising from its concise mathematics and easily obtained input data, the AHP has been of great interest to researchers of many difficult fields (Triantaphyllou and Mann, 1995). Raju et al. (2006) searched for the best compromise irrigation planning strategy for case study of Jayakwadi irrigation project of Mahahrastra, India. To rank the strategies linear programming, non-dominated irrigation planning strategies, Kohonen neural network, and multi-criteria analysis namely compromise programming have been applied. Anane et al. (2008) have located and ranked suitable sites for soil aquifer treatment in Jerba Island by integrating a single-objective AHP method into a GIS model. Sinha et al. (2008) have done a multi-parametric approach using AHP and integrates geo-morphological, land cover, topographic and social (population density) parameters to propose a Flood Risk Index for Kosi river. Kevin et al. (2009) has initiated AHP for finding of best management practices in selection and design of storm water. Park (2010) has compared the hydrological characteristics in several river basins and methodologies by using a GIS based distributed runoff model and AHP for the analysis of river basins based on their regional hydrological characteristics and considering their temporally and spatially-distributed physical properties. Tsiko and Haile (2011) integrated fuzzy logic and AHP to find the candidate water reservoir site of a Eritrean district. Machhiwal et al. (2011) have delineated ground water potential zones using integrated remote sensing, GIS, AHP and Multi Criterion Decision Making (MCDM) techniques in Udaipur district of Rajasthan, India.

The conceptual NAM model is also very popular in rainfall-runoff modeling and thereby very useful for setting the flood forecasting models. Dharmasena (1997) successfully applied MIKE-11 package to simulate one-dimensional unsteady flow. He also found conceptual models giving better results especially for rivers subjected to prolonged droughts. Tingsanchali and Gautam (2000) compared two lumped conceptual hydrologic models like tank and NAM with a neural network model applied in two river basins in Thailand. The works of Rabuffetti and Barbero (2005) also showed the application of NAM model. Keskin (2008) has applied MIKE-11 hydrological model coupled with two different numerical Weather Prediction Models to simulate runoff from precipitation in a semi-distributed manner in Filyos basin of Turkey. Bao et al. (2011) have applied NAM model for simulating a rainfall-runoff model within Kaidu river basin of China after estimating the areal rainfall through self- similarity topography method.

In this study different possible key RG networks have been designed from the available RG network in the Tel sub-catchment, Mahanadi, India using Hall method, Analytical Hierarchy Process (AHP), Hierarchical Clustering (HC) and Self Organisation Map (SOM). Efficiency of the key networks is tested by ANN, Fuzzy and NAM and the best network has been used for flood forecasting. Further, flood forecasting has been carried out with the key RG networks. Although, the best RG network has shown highest efficiency, simultaneously other networks were also tested with certain designated efficiency in order to use them at the time of failure of the best RG network.

6.3 STUDY AREA

A key raingauge network has been designed for Tel catchment upto Kantamal (Catchment area=19600 sq Km.). The downstream part below Hirakud having a catchment of around 48,500 sq. km. contributes substantially to flood at delta (9500 km^2) and is devoid of a sound flood forecasting system. This part has three main tributaries like Jeera, Ong and Tel with catchments 2383, 5128 and 22000 sq. km. respectively. Therefore, the contribution from the Tel catchment always remains predominant. Even the flood of 2008 is mainly due to the contribution of this tributary. It has produced a peak discharge of 33762 cumecs during 2008. Keeping this in view, establishment of a flood forecasting model at Kantamal upstream of Patharla is attempted in this study. The river Tel joins at Patharla to the main river Mahanadi at downstream of Hirakud reservoir. The location of Mahanadi basin with Kantamal sub-basin is shown in Fig.6.1.

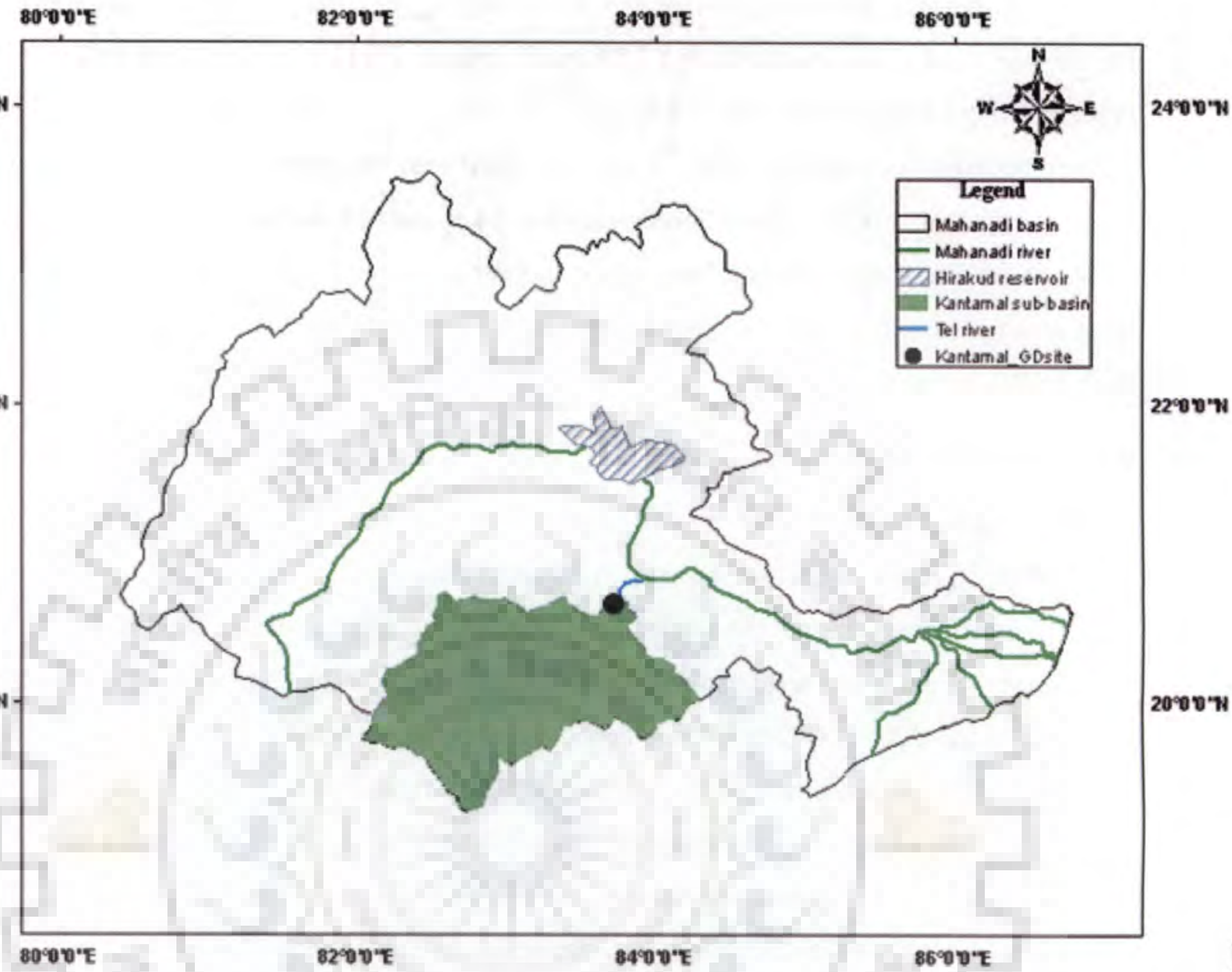


Fig. 6.1 Location of Mahanadi basin with Kantamal sub-basin

6.4 DATA USED

The daily rainfall data of 14 RGs of the study area have been collected for 6 years (2000-2005). The individual RGs have their designated IDs as fixed by concerned department (Department of Water Resources, Government of Orissa, India) shown in Table 6.1. The corresponding daily discharge at Kantamal and daily evaporation data of nearby station at Dasapalla are also collected. The available data are divided into calibration period (2000-2003) and validation period (2004 - 2005) and only monsoon period (June to October) is considered for the purpose. The physical characteristics of individual Thiessen polygon areas are derived from freely downloadable SRTM data of 90m resolution by using ARCGIS 9.3 software. The hydro-meteorological variables are collected from Department of Water Resources, Government of

Orissa. All the data and RG characteristics are normalized to 0 to 1 scale in order to have uniformity.

Table 6.1 Location of RG stations of Kantamal sub-basin with station IDs

Station ID	Location	Latitude(deg)	Longitude(deg)
KI8	Bhaskel	19.70833	82.13333
KI9	Kurumuli	19.25556	82.82667
M25	Sagada	20.64639	84.00056
M22	Magurbeda	20.78194	83.35833
M16	Goria	20.60556	83.57389
M1	Patora	20.66750	82.44111
M14	Baragaon	20.41111	83.21944
M19	Ichhapur	20.59583	82.59361
M15	Takala	20.25139	82.85222
M18	Chhatikud	19.97222	83.30278
M17A	Burat	20.18694	83.50722
M17	Tulaghat	20.27389	83.57389
M20	Surubali	20.17167	83.77944
R5	Pipalpankha	19.82667	84.33194

6.5 METHODOLOGY

In this study our basic aim is to find out key network of RGs (instead of taking information from all) that can be used for making reasonably accurate flood forecasts particularly during the time of distress (when the rainfall data of all the stations are not available due to various regions). The methodology is basically divided into five parts as:-

i) Derivation of storms in different RGs for applying Hall's method.
ii) Derivation of attributes (variables) for application into clustering methods and AHP.
iii) Process to find important (key) gauge networks using prioritizing methods Hall and AHP.
iv) Investigation of possible clusters influencing the model most, using HC and SOM.
v) Test the efficiency of key networks using ANN, Fuzzy and NAM model.
vi) Flood forecasting based on efficient network using ANN and Fuzzy.

The step of methodology is shown in Fig. 6.2.

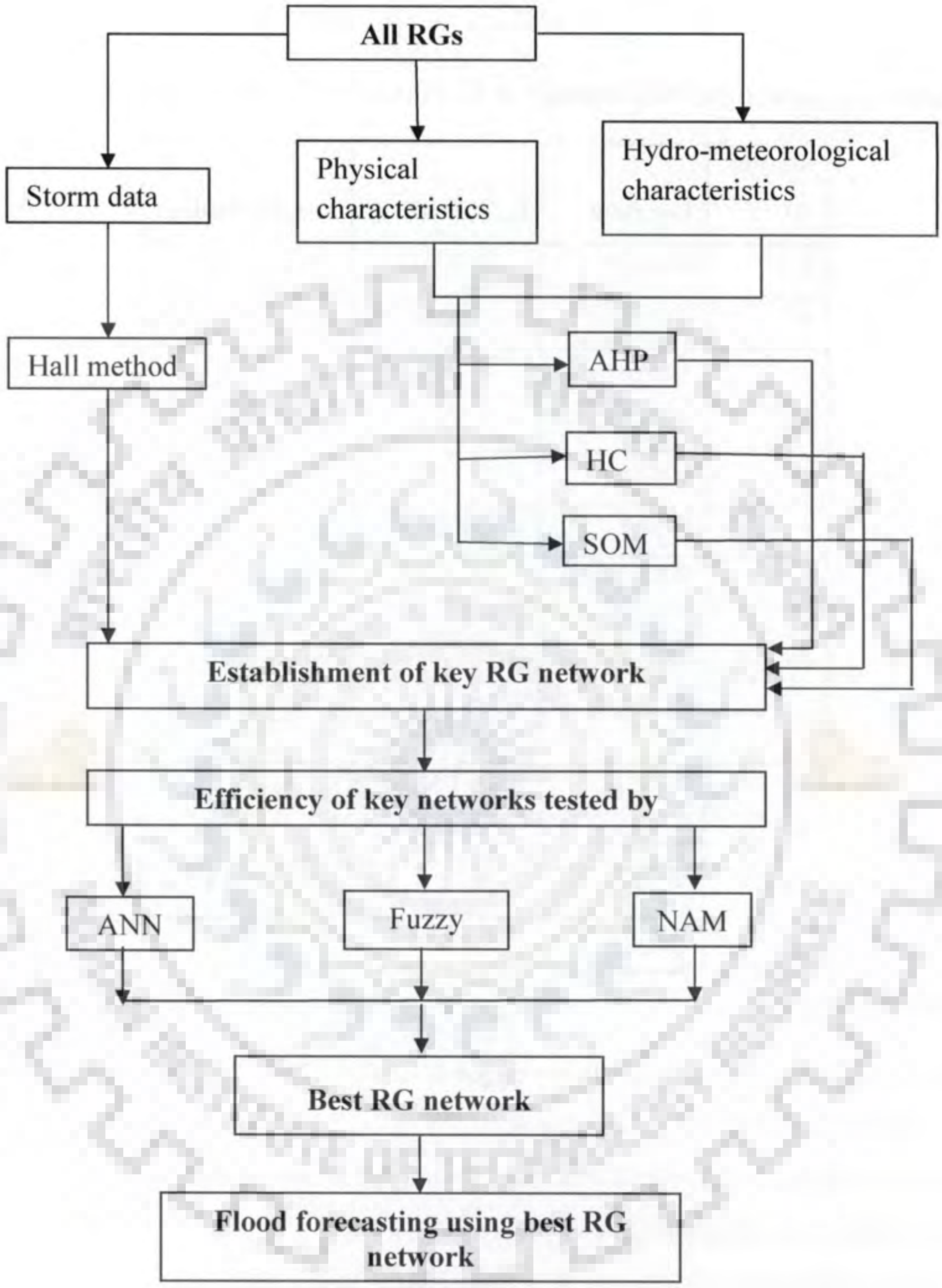


Fig. 6.2 Flow chart of methodology adopted in this study

The flow generation characteristic of each RG is different from each other. The property of each Thiessen area is represented through the physical and hydro-meteorological variables. The variables should be carefully selected and derived as these will be applied for finding key RG networks. In the process to find key RGs two prioritizing methods like Hall, AHP and two clustering methods SOM and HC are adopted. The Hall method forms the network of key RGs considering the storm characteristics only. Whereas, other methods like AHP, SOM, HC are dependent upon the characteristics of the Thiessen area occupied by each RG. The SOM is an unsupervised clustering technique, which is helpful in getting the possible number of divisions. The dendrogram generated by HC method also gives a justification to the decision of SOM. But AHP in this regard makes a ranking of RGs depending upon the influential characteristics of each Thiessen area. Thus key network decision is taken in a multifaceted way and efficiency of each is tested with both soft computing models (ANN, Fuzzy) and conceptual base model (NAM). The details of various steps of the methodology are given as follows:

6.5.1 Derivation of Storms

The isolated continuous rainfalls recorded at different RGs are selected as storms over a particular area. These storms are collected over a period, to study the potential RGs susceptible to high rainfall. Prioritization of RGs is made by applying Hall's method.

6.5.2 Derivation of Attributes (Variables)

The catchment is divided into Thiessen polygons corresponding to the existing RGs and each Thiessen polygon area is considered as one unit. The discharge produced by each Thiessen is dependent upon its hydro-meteorological and physical characteristics. Normally those characteristics are to be considered which can be very influential and achievable with less effort.

In this study daily average precipitation (AP) and maximum 1-day rainfall (1D) are used as hydro-meteorological variables. Physical attributes like Thiessen weight (TW), longest stream of Thiessen area (LS), average slope (SL) and drainage density (DD) are used for identification of key network. The average slope of the watershed has been derived from the DEM of the study area. Drainage density (DD) is derived by dividing the sum of total stream length within a catchment by the catchment area.

6.5.3 Procedures to Find Key RG network

6.5.3.1 Hall method

Hall (1972) suggested a rational method for determination of key station network using the equation

$$P_{av} = C + A_1P_1 + A_2P_2 + A_3P_3 + A_4P_4 \dots\dots\dots\dots\dots\dots\dots\dots + A_nP_n \quad (6.1)$$

Where,

P_{av} is the rainfall to be estimated from observed recorded rainfall at selected stations $P_1, P_2, P_3 .... P_n$.

$A_1, A_2, A_3 .... A_n$. are regression coefficients, C being a constant known intercept.

In order to establish key station network correlation coefficients between average of the storm rainfall and individual station rainfall are computed. The correlation coefficients thus obtained are arranged in a descending order and the highest correlation coefficient station is the first key station and its data is removed from the set. The next set is chosen in the same way and highest correlation coefficient bearer being the second key station and the process continues.

A key station network is investigated by computing the multiple correlations co-efficient of individual stations with that of average storm of the group. The stations are added one by one to the key station network and the total amount of variance at that stage is determined. With the addition of a RG to a network the multiple correlation coefficient increases and sum of the squares of deviation decreases till a stage is reached when improvement in either the multiple correlation coefficient or reduction in the sum of the squares of deviation will be negligible. The corresponding number of RGs at this stage is taken as the representative network for the purpose of determining aerial estimate of rainfall.

6.5.3.2 AHP

The Analytic Hierarchy Process (AHP) is a theory of measurement through pairwise comparisons and relies on the judgment of the expert to derive priority scales (Saaty, 2008). In AHP a goal is achieved in following four steps,

- Pairwise comparison of alternatives
- Extraction of priority vectors
- Finding consistency of pairwise judgments
- Ranking the priority alternatives

Pairwise comparison of alternatives

A matrix is framed containing the criteria/alternative with different choices. These alternatives are pre-selected for testing for a particular type of problem. This matrix remains the basis for evaluating different alternatives for achieving various selection criteria. In our study, matrix between RGs and property of Thiessen area is to be framed. Through this matrix both different choices of RGs as well as choices of different criteria/alternative properties of Thiessen area are compared. This comparison of matrices is dynamic and can be adjusted on separate applications.

A pairwise comparison matrix 'A' has to be formed, where the number in i^{th} row and j^{th} column gives the relative importance of O_i (objective) as compared with O_j (objective). The AHP conditional rules are shown in Table 6.2.

Table 6.2 AHP rules

a_{ij} (conditional numbers) =	**Conditional rules**
1	Two objectives are equal in importance.
3	If O_i is moderately more important than O_j .
5	If O_i is strongly more important than O_j .
7	If O_i is very strongly more important than O_j .
9	If O_i is extremely more important than O_j .
1.1 to 1.9	If O_i and O_j are very close.
2,4,6,8	Intermediate judgment values
Reciprocals like 1/3,1/5	O_j is moderately more important than O_i and so on for other reciprocals.

Extraction of priority vectors

The right eigen vector is decided for both choices (RGs) and alternatives (properties of Thiessen area). It is a collection of RGs ranking vectors (one for each alternative) and a single vector that ranks the alternatives.

Finding consistency of pairwise judgments

In order to prove the strength of the assumption consistency checks are applied. Whatever assumptions are taken for alternatives initially that must hold good for all choices otherwise it should be modified. The Consistency Index (CI) is as per following equations.

$$CI = (\lambda_{\max} - n)/(n-1) \tag{6.2}$$

Where, $\lambda_{\max}$ is the principal eigen value and n is the total number of activities. The consistency ratio (CR) is the ratio of Consistency Index to Random Index (RI) as follows:

$$CR = CI / RI \tag{6.3}$$

The value of RI is dependent on the size of matrix (Table 6.3). The value of CR should be within 0.1 for allowing the assumptions.

Table 6.3 Random Index (RI) values for different n

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0.0	0.0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.53	1.56	1.57	1.59

Ranking the priority alternatives

In this two level decision matrix, the priority vector for criteria is decided. The priority vectors are arranged in the matrix shown below. Simultaneously, the choices are again to be put in the same process of selection and priority vectors of choices are to be made against each criterion. Therefore, again a matrix of priority of vectors of choices is made as shown in eq. 6.4. For final evaluation of each choice the matrix multiplication of transposed matrix of priority vector of criteria are done with priority vector of choices (eq. 6.5). In our study, there are 6 criteria (alternatives) and each criterion has 14 choices (eq. 6.6). The final ranking of choices is presented in equations 6.7-6.9.

$$\begin{Bmatrix} A_1 \\ A_2 \\ . \\ A_6 \end{Bmatrix}^T \tag{6.4}$$

$$[A_1, A_2, \ldots\ldots\ldots A_6] \tag{6.5}$$

$$\begin{Bmatrix} a_{1,1}a_{2,1}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots a_{14,1} \\ b_{1,1}b_{2,1}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots b_{14,1} \\ . \\ . \\ n_{1,1}n_{2,1}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots n_{14,1} \end{Bmatrix} \tag{6.6}$$

Rank of choice, $a_1 = A_1^T a_{1,1} + A_2^T a_{2,1} \ldots\ldots\ldots\ldots + A_6^T a_{14,1}$ (6.7)

Rank of choice, $b_1 = A_1^T b_{1,1} + A_2^T b_{2,1} \ldots\ldots\ldots\ldots + A_6^T b_{14,1}$ (6.8)

Rank of choice, $n_1 = A_1^T n_{1,1} + A_2^T n_{2,1} \ldots\ldots\ldots\ldots + A_6^T n_{14,1}$ (6.9)

The overall consistency of the process can also be checked by using the eq. 6.10 with permissible value of $\overline{CR} \geq 0.1$. The overall consistency $\overline{(CR)}$ is

$$\overline{CR} = \sum_i w_i CI_i / \sum_i w_i RI_i \tag{6.10}$$

6.5.4 Application of Clustering Methods

Both Kohonen Self Organisation Map (SOM) and Hierachical Clustering are applied to the characteristics of each Thiessen occupied area in order to classify it. The details of these clustering methods are presented in Annexure-II.

6.5.5 Testing the Efficiency of Key Networks

6.5.5.1 ANN model

A 3-layered ANN model is selected to test the output of different models using MATLAB. The multi layered feed forward network is used with back propagation error modeling. The inputs are daily rainfall data of different RGs and evaporation data of a nearby site. The data are normalized to 0-1 scale. The optimum number of neurons in the hidden layer was identified using a trial and error procedure by varying the number of neurons in the hidden layer. The weights and biases of the networks were adjusted using gradient descent with momentum weight and bias learning function.

6.5.5.2 Fuzzy model

The TS fuzzy model using the subtractive clustering analysis has been attempted in the study. The data driven approach based on subtractive clustering has shown promising results in various hydrological modeling application (Lohani et al., 2005a, b). The purpose of subtractive clustering is to identify natural grouping of the data from a large dataset and finally to produce a concise representation of a system behavior (Lohani et al., 2006, 2007a).

The fuzzy model based on the assumption that the cluster estimation method when applied to a cluster of input and output data produces cluster centers where each cluster center represents a characteristic behavior of the system. Hence, each cluster center can be used as the basis of a rule that illustrates the system behavior. Thus the major parameter that needs to be identified in FIS model is the cluster radius. A smaller cluster radius will yield more clusters in a data and hence a greater number of rules. Simultaneously, it increases the model complexity and decreases the parsimony. The MATLAB software has been applied for solving this problem.

6.5.5.3 NAM

NAM is the abbreviation of the Danish "Nedbør-Afstrømnings-Model", meaning precipitation-runoff model, part of the rainfall-runoff (RR) module of the MIKE 11 river modeling system.

This model was originally developed by the Department of Hydrodynamics and Water Resources at the Technical University of Denmark. It is a deterministic, lumped and conceptual rainfall-runoff model which simulates the rainfall-runoff processes occurring at the catchment scale. This model is well adopted in different climatic zones of the world. A total of nine parameters being used as shown in Table 6.4.

Table 6.4 Parameters of NAM

Sl. No.	Parameters	Meaning
1	Umax	Maximum water content in surface storage
2	Lmax	Maximum water content in lower zone/root storage
3	CQOF	Overland flow coefficient
4	CKIF	Interflow drainage constant
5	TOF	Overland flow threshold
6	TIF	Interflow threshold
7	TG	Groundwater recharge threshold
8	CK1, CK2	Timing constant for overland flow, Timing constant for interflow
9	CKBF	Timing constant for base flow

NAM also uses auto calibration optimizing all 9 parameters automatically. The four different objectives like water balance, overall hydrograph shape, peak flows and low flows are used for auto calibration.

6.6 RESULTS AND DISCUSSIONS

The study area has 14 RGs and the corresponding discharge is measured at Kantamal G&D site. The evaporation data of nearby site named Dasapalla are used in the study. For the selection of key network stations Hall, AHP, SOM and HC methods have been applied. The results of these methods are discussed below:

6.6.1 Hall method

From the daily rainfall data of five years, total 29 storms are identified. The correlation coefficient between average of the storm rainfall and individual station rainfall are computed and arranged in a descending order (Table 6.5). The multiple correlation coefficients are shown in Table 6.6.

Table 6.5 Key RGs as per priority (Hall Method)

Normal sequence	Prioritized Stations	Correlations
R5	M1	0.901
KI8	M17A	0.876
KI9	M22	0.884
M25	M15	0.872
M22	M18	0.865
M16	M17	0.872
M1	M20	0.818
M14	M16	0.806
M19	M25	0.791
M15	M14	0.782
M18	M19	0.791
M17A	KI8	0.802
M17	KI9	0.882
M20	R5	0.491

A plot of MCC and RGs is shown in Fig. 6.3a. It is depicted from Fig. 6.3(a) that, the MCC is rising significantly upto addition of M17 but after the addition of M20 the MCC drops and then again it increases with the addition of M16. So M20 is a wrong fit here and after removing it the graph shows continuous growth (Fig.6.3b). The growth is not substantial after addition of M16. Thus a key network with 7 stations (M1, M17A, M22, M15, M18, M17 and M16) is finalized with MCC of 0.996 (Network HM).

Table 6.6 Multiple correlation coefficients of individual RGs with average storms of the sub-basin

	M1	M17A	M22	M15	M18	M17	M20	M16	M25	M14	M19	KI8	KI9	R5	Intercept	MCC
1	0.843														37.34	0.949
2	0.492	0.286													32.85	0.971
3	0.226	0.256	0.392												25.91	0.985
4	0.104	0.212	0.383	0.223											24.48	0.988
5	0.077	0.163	0.301	0.226	0.162										17.52	0.992
6	0.437	0.119	0.259	0.174	0.186	0.085									17.63	0.994
7	0.043	0.122	0.271	0.177	0.184	0.084	-0.015								17.90	0.993
8	0.137	0.098	0.353	0.076	0.077	0.063	-0.073	0.347							15.01	0.998
9	0.127	0.116	0.224	0.131	0.092	0.059	-0.075	0.236	0.135						15.40	0.998
10	0.171	0.117	0.191	0.079	0.045	0.068	-0.053	0.192	0.114	0.078					15.40	0.998
11	0.163	0.120	0.183	0.077	0.044	0.066	0.050	0.180	0.119	0.081	0.017				15.74	0.998
12	0.087	0.019	0.134	0.104	0.083	0.096	0.077	0.154	0.087	0.069	-0.015	0.122			6.00	0.999
13	0.065	0.064	0.112	0.077	0.059	0.077	0.049	0.139	0.069	0.065	0.071	0.091	0.059		5.72	0.999
14	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0	1.00

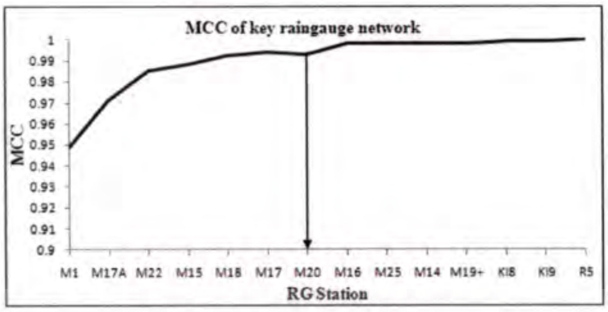


Fig. 6.3(a) Increase in MCC for formation of key RGs network (all stations)

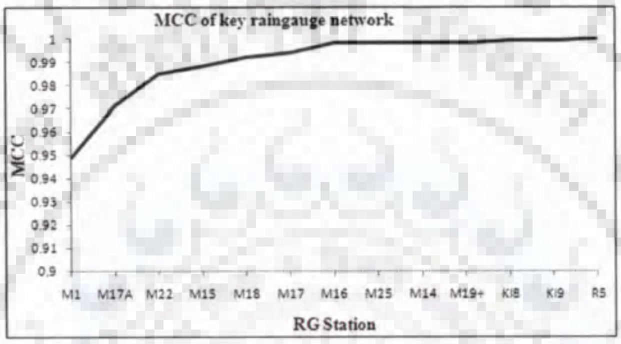


Fig.6.3(b) Increase in MCC for formation of key RGs network (after removing M20)

6.6.2 Analytic Hierarchy Process (AHP)

AHP has been applied according to Fig. 6.4. In order to form a key network of RGs, the first level (Criteria level) is to be carefully selected. Here, in first level 6 criteria viz. drainage density (DD), daily maximum rainfall (1D), daily average precipitation (AP), average slope of each Thiessen area (SL), Thiessen weight of each raingauge influenced area (TW) and longest stream of Thiessen area (LS) have been considered. In the second level there are choices of 14 RGs for each criterion.

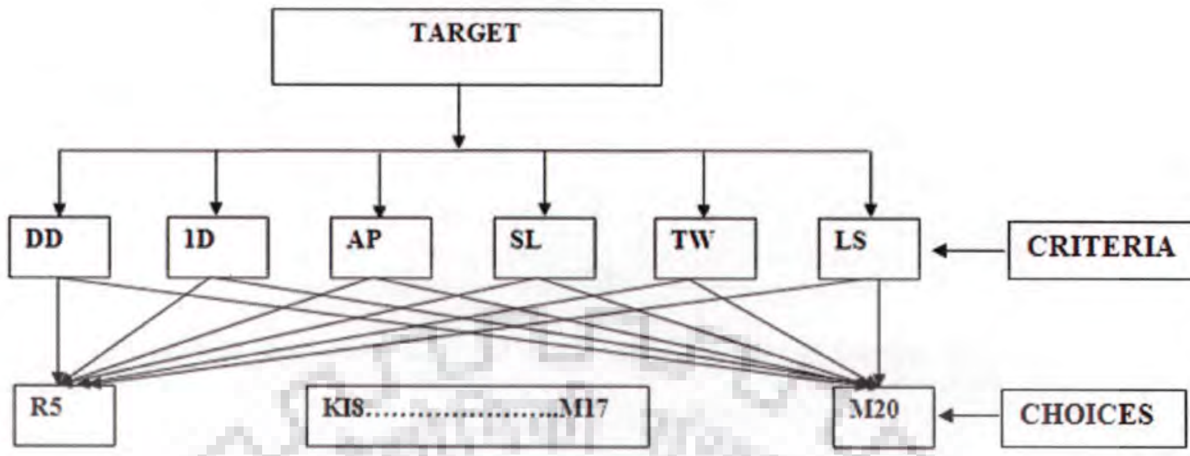


Fig. 6.4 AHP diagram of our study

For the application of this method the DEM of the study area was prepared from the freely available SRTM data. Thiessen areas of all 14 RGs were computed from the Thiessen polygon map of the study area (Fig.6.5). The hydro-meteorological characteristics like 1D and AP have been derived from the daily rainfall data. The physical characteristics like DD, SL, TW, LS have been obtained using ARCGIS. Further, on the basis of the characteristics of individual RGs AHP has been applied.

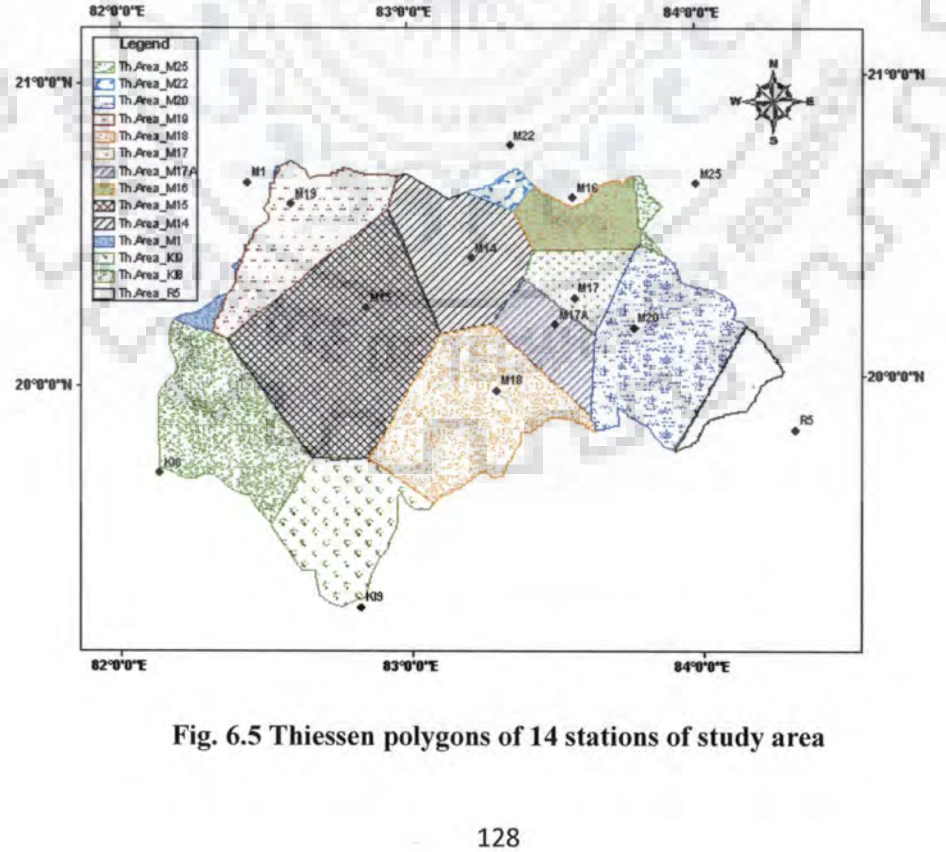


Fig. 6.5 Thiessen polygons of 14 stations of study area

In order to further strengthen the judgment, a principal component analysis has been applied over the six available criteria. It is seen from Table 6.7 that considering 3 PCs more than 90% variance is restored. The loadings of the first 3 PCs (having eigen value more than 1) in Table 6.8 say the first PC is governed by DD and RL, second by AP and 1D and third by SL. Based on principal component analysis and physical importance of various characteristics a judgment matrix on a scale of 1-9 was proposed. The same is given in (Table 6.9) keeping in the view of AHP rules of Table 6.2.

Table 6.7 Result of PCA

Principal Components	Eigen value	Percentage variance
1	2.587	43.121
2	1.602	26.692
3	1.264	21.070
4	0.266	4.431
5	0.188	3.140
6	0.093	1.546

Table 6.8 Loadings of first 3 PCs

Choice	PC1	PC2	PC3
SL	-0.300	0.3482	**-0.6244**
ID	0.3516	**0.5635**	0.2085
DD	**0.5192**	-0.0948	0.3757
AP	0.2363	**0.6701**	-0.0991
TW	0.4447	-0.2859	-0.4915
LS	**0.5128**	-0.1471	-0.4173

Table 6.9 Judgment matrix for criteria

Choice	LS	TW	SL	ID	DD	AP	
LS	1.0	0.333	0.2	0.2	0.11	0.2	
TW	3.0	1.0	0.333	0.333	0.20	0.5	
SL	4.0	3.0	1.0	0.333	0.333	0.5	
ID	5.0	3.0	3.0	1.0	0.5	2.0	
DD	9.0	5.0	3.0	2.0	1.0	2.0	
AP	5.0	2.0	2.0	0.5	0.5	1.0	
Sum	27.00	14.33	9.53	4.36	2.64	6.2	Total=64.07

The judgment matrix in Table 6.9 is then normalized by dividing each value with sum of that column and the principal eigen vector (priority vector) is obtained by adding the normalized values along each row (Table 6.10). Physical influence of each variable towards runoff generation is given importance for establishing the comparison matrix. From Table 6.10,Column 9, it is seen that DD has given maximum value of 34.33 followed by 1D=22.63, AP=17.17, SL=14.31, TW=8.37 and LS=3.19, which show that DD is highly influential criterion, followed by 1D, AP, SL, TW and LS.

Table 6.10 Normalized judgment matrix

Choice	LS	TW	SL	ID	DD	AP	Sum	Priority vector (%)
LS	1.561	0.520	0.312	0.312	0.173	0.312	3.191	**3.19**
TW	4.682	1.561	0.520	0.520	0.312	0.780	8.375	**8.37**
SL	6.242	4.682	1.561	0.520	0.520	0.780	14.306	**14.31**
1D	7.803	4.682	4.682	1.561	0.780	3.121	22.629	**22.63**
DD	14.045	7.803	4.682	3.121	1.561	3.121	34.333	**34.33**
AP	7.803	3.121	3.121	0.780	0.780	1.561	17.167	**17.17**
Sum	**42.136**	**22.369**	**14.878**	**6.815**	**4.127**	**9.676**	**100.000**	**100.000**

The pairwise comparison of initial assumptions of the physical and hydro-meteorological criteria has been checked for its consistency. The principal eigen value $\lambda_{\max}$ which is the sum of the product of sum of each column of judgment matrix and its corresponding priority vector comes as 6.386 and $CI = 0.077$. So $CR = 0.077/1.24$, which comes as $0.062 \leq 0.1$. So the assumptions are consistent. The value of RI is obtained as 1.24 for $n = 6$ from Table 6.3.

The same steps again continued for second level of hierarchy. Here there are 14 choices (RG) against 6 criteria (DD, 1D, AP, SL, TW and LS). So ranking of choices of each RG are done against each criterion. For putting 14 RGs in a 1-9 scale we grouped them in to 10 classes and judge each of them as per their rank with respect to individual criterion like DD, 1D, AP, SL, TW and LS. Thus 6 such priority vectors according to first level of hierarchy are made available for individual choices. Then the final priority vectors for finding best RGs are obtained by multiplying the value of each choice with corresponding weights of criteria obtained from Table 6.11.

Table 6.11 Overall ranking of choices

↓ Choices / Weights →	DD	1D	AP	SL	TW	LS	Priority vector (%)
	0.3433	**0.2263**	**0.1717**	**0.1431**	**0.0837**	**0.0319**	
R5	1.77	1.77	2.19	20.38	2.06	2.19	4.54
KI8	14.19	2.69	4.39	3.11	11.15	8.54	7.89
KI9	11.10	17.04	17.09	3.11	6.21	4.38	11.70
M25	1.76	2.17	1.78	14.08	1.78	1.77	3.62
M22	4.36	8.54	3.10	1.79	2.19	2.19	4.47
M16	20.16	2.19	1.78	2.20	3.36	3.09	8.42
M1	3.07	6.20	6.22	4.41	1.78	1.77	4.36
M14	6.18	4.38	4.39	1.87	8.28	13.98	5.27
M19	2.05	2.19	2.19	4.41	4.38	4.38	2.71
M15	8.51	3.09	3.09	2.20	20.25	20.24	6.81
M18	3.34	11.14	11.17	11.22	17.06	6.20	8.82
M17A	16.97	20.24	14.02	6.24	4.38	11.14	14.43
M17	2.18	4.38	20.02	7.80	3.13	3.09	6.65
M20	4.36	13.98	8.57	17.16	13.99	17.04	10.30

Thus each RG is ranked as per the characteristic it bear with respect to six criteria fixed for each RG. An overall consistency is again checked for justifying all the assumptions made earlier. The

$\overline{CR}$ value obtained using the equation 6.10 is 0.0569 which is ≤ 0.1. On the basis of the final priority vector ranking of individual RGs has been made and shown in Table 6.12 against that of Hall method.

Table 6.12 Prioritisation of RGs as per Hall's method and AHP

Sl. No.	Station name (normal sequence)	Prioritized Stations	
		Hall method	AHP
1	R5	M1	M17A
2	KI8	M17A	KI9
3	KI9	M22	M20
4	M25	M15	M18
5	M22	M18	M16
6	M16	M17	KI8
7	M1	M20	M15
8	M14	M16	M17
9	M19	M25	M14
10	M15	M14	R5
11	M18	M19	M22
12	M17A	KI8	M1
13	M17	KI9	M25
14	M20	R5	M19

6.6.3 Self Organization Map (SOM)

Further another unsupervised method Kohonen Self Organization Map is trialed with same datasets in order to find the possible number of groups and RGs constituting each group. The NNclust software (*http://www.geocities.com/adotsaha/NNSOMinEXCEL.html.*) has been used in this regard. Several combination of learning parameter, sigma for Gaussian neighborhood and training cycles are considered. The network gave consistent result at learning parameter 0.35 to 0.1, Gaussian neighborhood 30% to 1% and training cycle of 200. The results (Fig.6.6) illustrates that there are two possible clusters separating 14 RGs into two groups of 7 RGs in each.

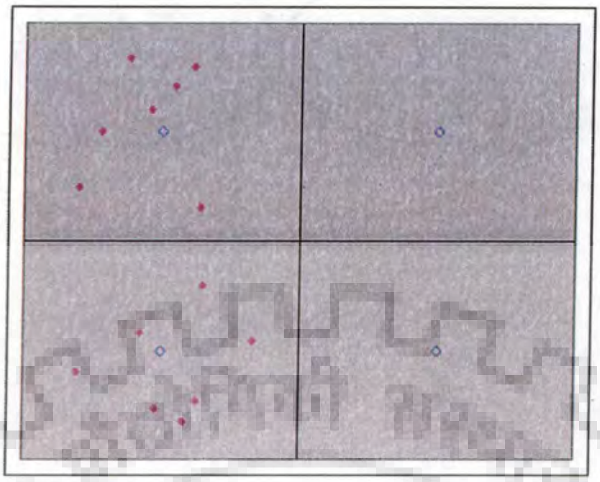


Fig. 6.6 RGs distributed in two clusters through Kohonen Self Organization Map

Allotments of RGs in two different clusters are shown in Table 6.13. Taking each group as a possible network, two possible networks have been obtained with KI8, KI9, M14, M15, M18, M17A, M20 as SOM1 and R5, M25, M22, M16, M1, M19, M17 as SOM2.

Table 6.13 Cluster assignment as per Kohonen Self Organization Map

Observation ID	R5	KI8	KI9	M25	M22	M16	M1	M14	M19	M15	M18	M17A	M17	M20
Cluster ID	2	1	1	2	2	2	2	1	2	1	1	1	2	1

6.6.4 Hierarchical Clustering

The same variables of 14 RGs have been considered for hierarchical clustering. Dendogram was obtained using software *PAST (version 1.62)*. It is depicted from the dendrogram (Fig.6.7) that all the available 14 RGs can be separated into 2 clusters. Further each cluster has been taken as one model (HC1, HC2) in our study. The results of SOM and HC are same.

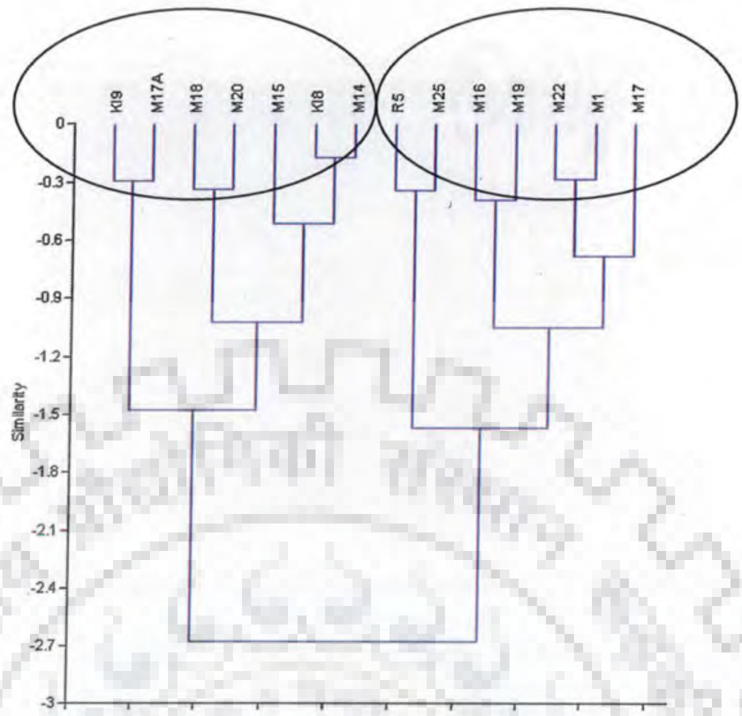


Fig. 6.7 The dendrogram of HC dividing 14 RGs in two groups

6.6.5 Performance Measure of Key RG Networks

Two prioritising methods (Hall and AHP) and two clustering methods (HC and SOM) have been applied in this study. The key RG networks from AHP and Hall methods are tested using ANN and Fuzzy techniques. The performance criteria are fixed as R^2 (coefficient of determination), efficiency (Nash-Sutcliffe criterion) and RMSE (m^3/s) are tested for both networks adding one station each time. It is observed that a major portion of performance is achieved with 7 RGs in AHP network. The R^2, RMSE and efficiency of AHP network with 7 RGs are 0.9319, 441.107 and 0.8618 and that of Hall method with same 7 RGs is 0.9021, 521.8355 and 0.8123 during validation. The result of AHP network shows better result than Hall network. The fuzzy logic applied performance measures remain higher than that of ANN results.

The two clustering methods HC and SOM also show presence of 7 RGs in each cluster. So for comparative study 7 best RGs are chosen from two network models and 2 each from clustering methods. Thus a total 6 best key network models (1 from Hall method, 1 from AHP, and 2 each from SOM and HC) have been considered and tested for their efficiency and are presented in Table 6.14.

Table 6.14 Key rain gauge networks by all four methods

Priority Methods	Networks	Network name	Rain gauge names
Hall	1	HM	M1,M17A,M22,M15,M18,M17,M16
AHP	1	AHP	M17A,KI9,M20,M18,M16,KI8,M15
SOM	1	SOM1	KI8,KI9,M14,M15,M18,M17A,M20
	2	SOM2	R5,M25,M22,M16,M1,M19,M17
HC	1	HC1	KI9, M17A, M18, M20,M15, KI8,M14
	2	HC2	R5,M25, M16, M19, M22, M1, M17

The models so chosen are tested for rainfall-runoff modeling using ANN, Fuzzy logic and conceptual NAM model. The input data contains 6 years daily rainfall, evaporation and discharge data divided into calibration and validation periods. The same performance criteria are fixed as applied earlier for ANN, Fuzzy and NAM models.

First of all a MLFF ANN network has been attempted with varying number of hidden layers, hidden neurons and iterations. Only one hidden layer has been fixed for all models in order to avoid the model complexity. For different models it has been observed that the hidden neurons vary from 1 to 10 in numbers, learning rate from 0.4 to 0.6, momentum constant from 0.7 to 0.9 and epochs from 50 to 150 with increment of 20. The best performance measures for different models are collected and shown in Table 6.15. For Fuzzy logic approach the Takagi–Sugeno algorithm (Lohani et al., 2007a) has been applied. The subtractive cluster radius has been varied from 0.1 to 0.7 with an increment of 0.05 in each trial. The performance measures are put in Table 6.15.

Table 6.15 Performance measure of different networks

Method	RG Network	Coeff. of determination(R^2) Calibration	Coeff. of determination(R^2) Validation	RMSE (m^3/s) Calibration	RMSE (m^3/s) Validation	Efficiency Calibration	Efficiency Validation
ANN model							
Hall	1	0.925	0.883	528.069	571.716	0.8556	0.774
AHP	1	0.946	0.893	486.983	570.617	0.844	0.774
HC	1	0.9251	0.862	567.534	615.797	0.856	0.738
	2	0.892	0.693	675.819	875.795	0.795	0.471
SOM	1	0.9251	0.862	567.534	615.797	0.856	0.738
	2	0.8920	0.6931	675.819	875.795	0.795	0.471
Fuzzy model							
Hall	1	0.903	0.902	539.944	521.835	0.816	0.812
AHP	1	0.9243	0.9319	470.461	441.107	0.8544	0.8618
HC	1	0.904	0.9054	537.938	511.780	0.8174	0.819
	2	0.841	0.7396	808.815	815.188	0.7073	0.542
SOM	1	0.904	0.905	537.938	511.780	0.8174	0.819
	2	0.841	0.739	808.815	815.188	0.7073	0.542
NAM							
Hall	1	0.753	0.611	884.135	893.195	0.650	0.541
AHP	1	0.830	0.602	616.477	754.757	0.830	0.601
HC	1	0.744	0.441	756.87	892.64	0.744	0.442
	2	0.614	0.546	928.06	835.95	0.615	0.511
SOM	1	0.744	0.441	756.87	892.64	0.744	0.442
	2	0.614	0.546	928.06	835.95	0.615	0.511

Similarly in case of NAM model the initial parameters have been derived by auto-calibration. Further, these parameters have been fine tuned using the available basin information. The parameters derived for best calibration results have been used for the validation series. Finally, the performance measures have been derived in the same way as in the case of ANN and Fuzzy approach and presented in Table 6.15.

It is revealed from the analysis that RG networks established by AHP perform better for all the models derived using ANN, Fuzzy and NAM. However the RG network established by Hall method also performs better but remains inferior to AHP. When comparing the results obtained from all the models (Table 6.13), the AHP network with Fuzzy R-R model has been observed as the best model with 86.18% efficiency in validation. Similarly, the same AHP network provides the lowest RMSE of 441.107m^3/s.

6.6.6 Best Networks

The best combination has been obtained as AHP which gives coefficient of determination R^2 in calibration and validation as 0.946 and 0.893, RMSE 486.983, 570.6169 and efficiency as 0.8439, 0.7743 as per ANN. While, the AHP derived fuzzy R-R model gives coefficient of determination R^2 0.9243, 0.9319, RMSE 470.461, 441.107, efficiency 0.8544, 0.8618 and the NAM model gives coefficient of determination R^2 0.830, 0.602, RMSE 616.4774, 754.7571 and efficiency as 0.830 and 0.601. Comparing the performance criteria of other models the SOM1 or HC1 model is the second best and HM is the third best network model. So in absence of the first network second and third can be attempted to make a reasonable forecast. As per the performance criteria 3 best networks are detailed in Table 6.16.

Table 6.16 Best RGs networks

Network	Rank	RGs involved
AHP	Best	M17A,KI9,M20,M18,M16,KI8,M15
SOM1 OR HC1	Second	KI8,KI9,M14,M15,M18,M17A,M20
HM	Third	M1,M17A,M22,M15,M18,M17,M16

It is revealed from the Table 6.14 that the stations M17A, M18, M15 are common in 3 top networks whereas KI8, KI9, M20 are common to AHP and SOM1/HC1 network and M16 is common to HM and AHP network. Again AHP and SOM network has six and AHP and HM have 4 RGs common between them. Addition of M16 to the network of AHP makes it more efficient. The basic difference here between cluster methods and prioritizing methods is that cluster methods are limited to the number of sites contained in a cluster and it is difficult to find which one is most influential. But in prioritizing methods a sequence with priority is achieved. In this study both the clustering methods dividing the 14 RGs equally. The stations M17A, KI9, M20 are highly responsive as per storm and physiographic characteristics. The RGs selected according to AHP network establish a good rainfall-runoff model.

6.6.7 Flood Forecasting with Best RG Network

After finding the best 3 RG networks flood forecasting has been attempted at Kantamal gauge and discharge site using Fuzzy logic and ANN models. Taking these networks the inputs are fixed as discharge at Kantamal (Q_t) as per equation 6.14-6.16.

$$Q_{t(AHP)} = f\{M17A_{(t-1)}, M17A_{(t)}, KI9_{(t-3)}, KI9_{(t-2)}, M20_{(t-1)}, M20_{(t)}, M18_{(t-2)}, M18_{(t-1)}, M18_{(t)}, KI8_{(t-3)}, M15_{(t-1)}, M15_{(t)}, M16_{(t)}, Evaporation_{(t)}\}. \quad (6.14)$$

$$Q_{t(SOM1)} = f\{KI8_{(t-3)}, KI9_{(t-3)}, M14_{(t-1)}, M14_{(t)}, M15_{(t-1)}, M15_{(t)}, M18_{(t-2)}, M18_{(t-1)}, M18_{(t)}, M17A_{(t-1)}, M17_{(t)}, M20_{(t-1)} M20_{(t)}, Evaporation_{(t)}\}. \quad (6.15)$$

$$Q_{t(HM)} = f\{M1_{(t-1)}, M17A_{(t)}, M22_{(t-1)}, M15_{(t-1)}, M18_{(t-2)}, M18_{(t-1)}, M18_{(t)}, M17_{(t-1)}, M16_{(t)}, Evaporation_{(t)}\}. \quad (6.16)$$

The inputs are put into a 3 layered feed forward network starting from with 1 neuron in hidden layer to 10 and one output neuron. The attempt is made for flood forecasting at one and two day lead periods. Same way the Takagi Sugeno fuzzy model is also applied over same inputs with initial cluster radius of 0.1 goes upto 0.7. The results of the both models are shown in Table 6.17.

Table 6.17 Performance measures of models for 1-day lead period flood forecasting

Network	Model	Correlations (R^2)		RMSE(m^3/s)		Efficiency	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
AHP	ANN	0.9504	0.8775	465.1316	590.3663	0.9032	0.7595
	Fuzzy	0.9195	0.9105	587.6704	500.2123	0.8455	0.8274
SOM1	ANN	0.9478	0.8587	476.6766	620.5795	0.8984	0.7343
	Fuzzy	0.9116	0.8984	614.7763	530.0090	0.8309	0.8062
HM	ANN	0.9221	0.8471	578.3719	649.6842	0.8503	0.7090
	Fuzzy	0.8739	0.8582	726.7806	625.8077	0.7637	0.7300

However, the one day lead period results are shown here, because the two day lead period results are with poor efficiency, so eliminated from the study. The Fig. 6.8-6.10 shows the results of ANN and Fuzzy models over the observed floods. It is again verified that a Fuzzy logic based flood forecasting model gives higher efficiency than ANN model.

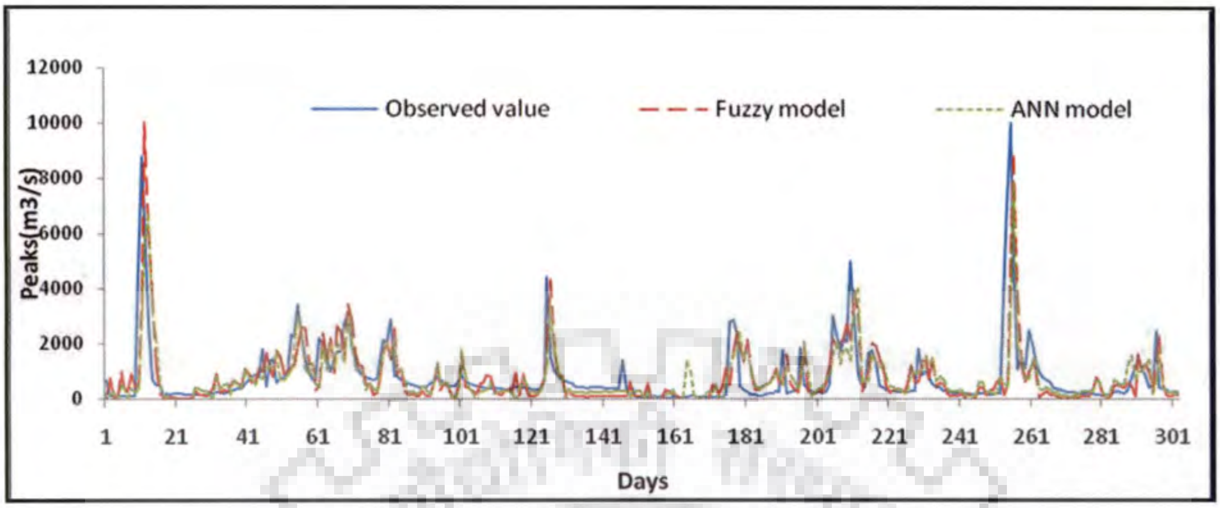


Fig. 6.8 Flood forecasting at Kantamal with 1-day lead time (AHP)

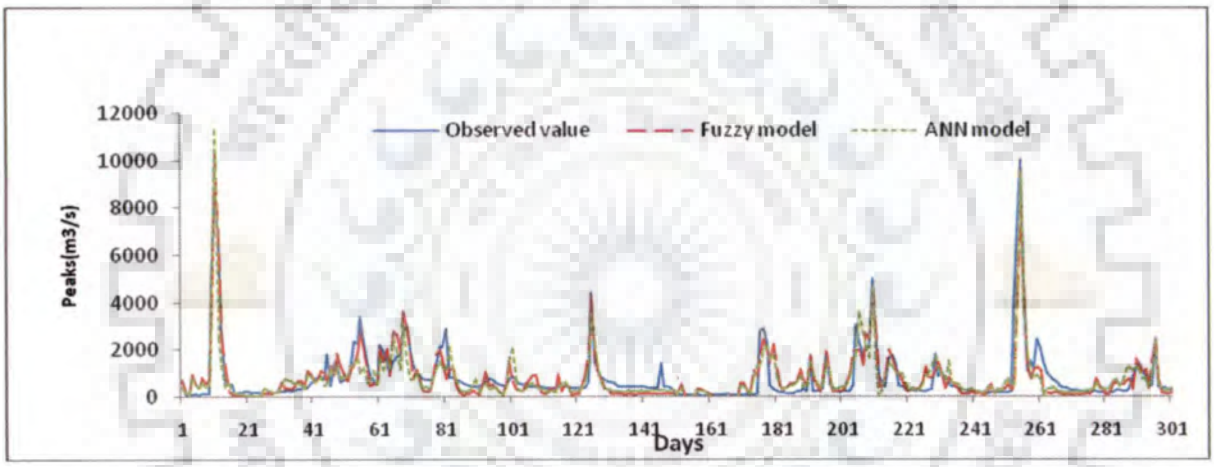


Fig. 6.9 Flood forecasting at Kantamal with 1-day lead time (SOM)

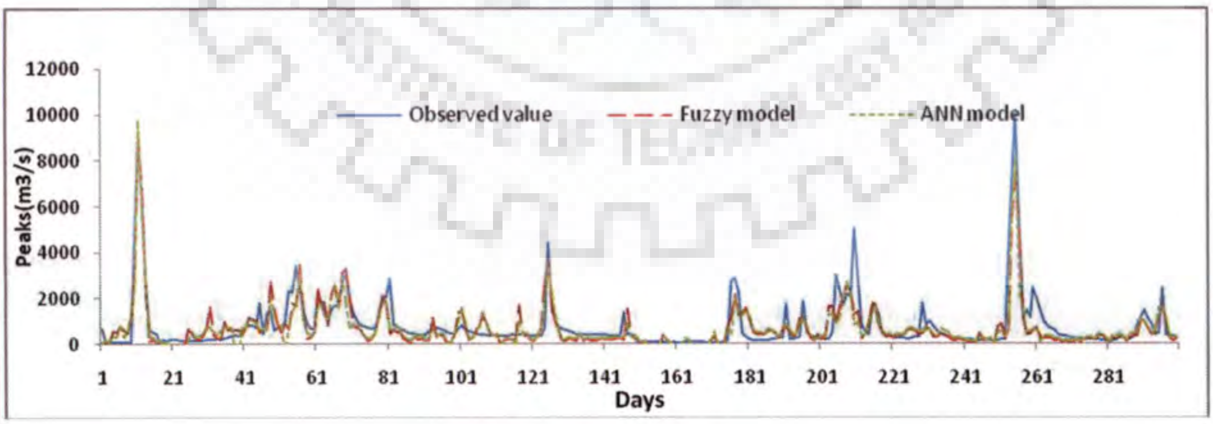


Fig. 6.10 Flood forecasting at Kantamal with 1-day lead time (HM)

6.7 CONCLUSIONS

Selection of key RG network is an important task in any reliable flood forecasting system. In this study, a procedure for the design of key RG network for flood forecasting is developed and demonstrated through a case study using the data of Kantamal sub-basin of lower Mahanadi. Two prioritization methods viz. Hall method and Analytic Hierarchy Process and two clustering methods like Self Organization Map and Hierachical Clustering are applied to form the key RG network. Prioritization of RGs as per Hall method is based on the storm characteristics, whereas AHP, HC, SOM give importance to both physical and hydro-meteorological characteristics. The following conclusions are drawn from this study:

1) Prioritisation methods are better than clustering methods as clustering methods may end up with unequal number of raingauge stations (one containing very high and other containing very less number of stations) in two groups. AHP method is better than other methods in finding the best RG network because it takes care of the merits of the each characteristic of RG occupied Thiessen areas.
2) Flood forecasting is possible to a reasonable efficiency, using the best key network consisting of only seven defined RGs instead of 14 established stations.
3) The Fuzzy logic based method is better as compared to ANN and NAM model, in rainfall-runoff modeling as well as flood forecasting.

6.8 SCOPE FOR FURTHER WORK

The study has a wide scope for further work in the area. The possible directions in which further work can be undertaken are listed below:

i) In this study the prioritized raingauge network has been established by using Analytic Hierarchy Process (AHP). The study may be carried out with more number of variables. The methodology may also be tested in other basins.
ii) The methodology of raingauge network design and its application in flood forecasting may be applied and tested in other basins located in different hydro-meteorological zones of the world.

CHAPTER 7 – CONCLUSIONS & SCOPE FOR FURTHER WORK

7.1 CONCLUSIONS

In the present study efforts have been made to develop regional flood formulae for the entire Mahanadi basin using L-moment and prioritized variables based approach. For the lower reach of Mahanadi basin (downstream of Hirakud dam) flood forecasting models have been developed using soft computing techniques like ANN and Fuzzy logic. The performance of soft computing models has been compared with conventional and conceptual models. Also by raingauge network has been developed for Kantamal sub-basin of lower Mahanadi basin and its performance in flood forecasting has been evaluated.

The conclusions drawn from this study are summarized in following sections:

7.1.1 Development of Regional Flood Formulae for Mahanadi basin

The study describes the division of Mahanadi basin into two homogeneous clusters for making flood frequency analysis using optimum number of prioritized variables. The study also displays how the prioritized variables influence the clustering process. During prioritization of variables using PCA both statistical and physical importance of the variables towards runoff generation are considered. The following conclusions have been drawn from this study:

(i) FCM clustering method was found to be robust.

(ii) The SOM and AP are helpful in deciding the number of clusters.

(iii) Reducing the dimensionality of variables by using 4 variables out of 7 available have not put any significant impact on homogeneity and cluster formation.

(iv) The Generalised Pareto (GP) distribution holds good for cluster-1 and it contains the areas which can contribute substantially towards runoff generation due to high slope and drainage density characteristics.

(v) The cluster-2 contains areas with low runoff generation capacity as compared to cluster-1. Generalised Extreme Value (GEV) is the robust distribution for this cluster.

7.1.2 Development of a flood forecasting model in the reach downstream of Hirakud

In a flood prone basin like Mahanadi controlling floods through structural measures is a difficult and costly task. The non-structural measures like flood forecasting is a better solution towards sufferings of eight coastal districts of Orissa. The Hirakud dam controls only 83000 km^2

catchment and nearly 58000 km^2 downstream area remains uncontrolled. Existing flood forecasting system is based on time lag approach. Thus development of a flood forecasting model using soft computing technique is expected to improve the existing forecast quite significantly. In this study both peak discharge as well as 3-hourly discharges at Mundali have been forecasted. The following conclusions are drawn from this study:

i) Statistical, ANN and Fuzzy logic models are applied for forecasting of flows at Barmul and Mundali. The performance of TS-fuzzy method is better than MLFF-ANN, RBF-ANN and statistical methods.

ii) The studies indicate that RBF performs better than MLP in forecasting of flood flows in lower reach of Mahanadi basin.

iii) For forecasting of peak flows 3 models viz. Khairmal-Mundali (KM), Barmul-Mundali (BM) and Khairmal-Barmul-Mundali (KBM) have been developed. Performance wise KBM is better than BM and BM is better than KM. However, lead time is maximum in case of KM and performance of KM is also satisfactory.

iv) For forecasting of 3-hourly flow data, three recommended models are KM-7, BM-3 and KBM. Although, KBM is the best model out of all 15 model combinations, KM-7 and BM-3 can be used as alternate models.

v) The lead time is maximum case of KM-7 model. Hence, Model KM-7 is recommended for the situations where lead time of 24-32 is required, even though its performance is slightly inferior to BM-3 and KBM models.

7.1.3 Development of a raingauge network for Kantamal sub-basin of lower Mahanadi basin for flood forecasting

Selection of key RG network is an important task in any reliable flood forecasting system. In this study, a procedure for the design of key RG network for flood forecasting is developed and demonstrated through a case study using the data of Kantamal sub-basin of lower Mahanadi. Two prioritization methods viz. Hall method and Analytic Hierarchy Process and two clustering methods like Self Organization Map and Hierachical Clustering are applied to form the key RG network. Prioritization of RGs as per Hall method is based on the storm characteristics, whereas AHP, HC, SOM give importance to both physical and hydro-meteorological characteristics. The following conclusions are drawn from this study:

i) Prioritisation methods are better than clustering methods as clustering methods may end up with unequal number of raingauge stations (one containing very high and other containing very less number of stations) in two groups. AHP method is better than other methods in finding the best RG network because it takes care of the merits of the each characteristic of RG occupied Thiessen areas.

ii) Flood forecasting is possible to a reasonable efficiency, using the best key network consisting of only seven defined RGs instead of 14 established stations.

iii) The Fuzzy logic based method is better as compared to ANN and NAM model, in rainfall-runoff modeling as well as flood forecasting.

7.2 SCOPE FOR FURTHER WORK

The study has a wide scope for further work in the area. The possible directions in which further work can be undertaken are listed below:

i) The regional flood frequency analysis has been carried out based on the prioritized variables and clustering methods. In the present study only few variables have been considered. The other variables influencing runoff may be considered in future studies.

ii) The result of regional flood frequency analysis may be improved further by including data of more number of stations analysis more length of data.

iii) The regional flood frequency may be extended using partial duration series approach.

iv) The flood forecasting is done in the downstream catchment using 3-hourly discharge data of 3 downstream stations of Hirakud reservoir. Non-availability of similar duration of rainfall, discharge of other tributaries and information of Hirakud release restricted the study to use the discharge data of 3 stations only. Further, flood forecasting studies may be taken up with the help of the forecasted rainfall to achieve a better lead time.

v) A hydro-meteorological data observation network should be established in the Mahanadi basin in order to develop a flood forecasting model for the entire Mahanadi basin for the better management of floods.

vi) There is always a conflict of superiority between RBF-ANN and MLFF-ANN network. Although in our study RBF has given a better performance, it requires further extensive testing in different watersheds.

vii) In this study the prioritized raingauge network has been established by using Analytic Hierarchy Process (AHP). The study may be carried out with more number of variables. The methodology may also be tested in other basins.

viii) The methodology of raingauge network design and its application in flood forecasting may be applied and tested in other basins located in different hydro-meteorological zones of the world.

REFERENCES

Abonyi, J., Babuska, R. and Szeifert, F. (2002) Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models, IEEE Transactions on Systems, Man and Cybernetics, Part B, 32(5), 612-621.

Acreman, M.C. and Sinclair, C.D. (1986) Classification of drainage basins according to their physical characteristics: An application for Flood Frequency Analysis in Scotland, Journal of Hydrology, 84(3-4), 365-380.

Ahmad, S. and Simonovic, S.P. (2005) An artificial neural network model for generating hydrograph from hydro-meteorological parameters. Journal of Hydrology, 315(1-4), 236-251.

Akaike H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.

Anders, U., Korn, O. (1999) Model selection in neural networks. Neural Networks, 12, 309–323.

Andrews, D.F. (1972) Plots of high dimensional data, Biometrics, Special multivariate issue, 28(1), 125-136.

Angelov, P. (2004) An approach for fuzzy rule base adaptation using on-line clustering. International Journal of Approximate Reasoning, 35(3), 275-289.

Anmala, J., Zhang, B. and Govindaraju, R.S. (2000) Comparison of ANNs and empirical approaches for predicting watershed runoff, Journal of Water Resources Planning and Management, 126(3), 156-166.

Aqil, M., Kita, Y., Yano, A. and Nishiyama, S. (2006) A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behavior of runoff. Journal of Hydrology, 337, 22-37.

ASCE Task Committee on Application of Artificial Neural Network in Hydrology (2000,a) Artificial neural network in hydrology. I: Preliminary concepts, Journal of Hydrologic Engineering, 5(2), 115-123.

ASCE Task Committee on Application of Artificial Neural Network in Hydrology (2000,b) Artificial neural network in hydrology. II: Hydrologic applications, Journal of Hydrologic Engineering, 5(2), 124-137.

Atiya, A., El-Shoura, S., Shaheen, I. and El-Sheriff, M. (1999) A comparison between neural network forecasting techniques – case study: River flow forecasting, IEEE transaction on Neural Network, 10(2), 402-409.

Awchi, T.A. (2008) Application of radial basis function neural networks for reference evapotranspiration prediction, Al-Rafidain Engineering, 16(1), 117-129.

Bao, A. M., Liu, H. L., Chen, X. and Pan, X. (2011) The effect of estimating areal rainfall using self-similarity topography method on simulation accuracy of runoff prediction, Hydrological Processes, DOI: 10.1002/hyp.8078.

Bardosy, A. (2006) Fuzzy sets in rainfall/runoff modeling, DOI 10.1002/0470848944.hsa137, Encyclopaedia of Hydrological Sciences, Wiley online library.

Berger, K.P. and Entekhabi, D. (2001) Basin hydrologic response relations to distributed physiographic descriptors and climate, Journal of Hydrology, 247(3-4), 169-182.

Beaulieu, C., Seidou, O., Ourada, T.B.M.J. and Zhang, X. (2009) Inter-comparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change point models, Water Resources Research, 45(8), 15.

Bezdek, J. C. (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York.

Bhatt, V.K. (2003) Estimation of extreme flows for ungauged catchments, Ph.D.Thesis, IIT, Roorkee.

Bhattacharjya, R.K., Datta, B. and Satish, M.G. (2007) Artificial neural networks approximation of density dependent salt water intrusion process in coastal aquifers. Journal of Hydrologic Engineering, 12(3), 273-282.

Bianchini, M., Frasconi, P. and Gori, M. (1995) Learning without local minima in radial basis function networks. IEEE Transactions on Neural Networks 6(3), 749-755.

Birikundavyi, S., Labib, R. Trung, H.T. and Rousselle, J. (2002) Performance of neural networks in daily streamflow forecasting, Journal of Hydrologic Engineering, 7(5), 392-398.

Bishop, C.M. (1994) Neural networks and their applications, Review of Scientific Instruments, 65, 1803-1832.

Bhattacharya, B. and Solomatine, D.P. (2005) Neural network and M5 model trees in modeling water level discharge relationship, Journal of Neurocomputing, 63, 381-396.

Blazkova, S. and Beven, K. (2004) Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic. Journal of Hydrology, 292(1-4), 153-172.

Borujeni, S.C., Sulaiman, W.N.A. and Eslamian, S. (2009) Regional flood frequency analysis using L-moments for North Karoon basin, Iran, Journal of Flood Engineering, 1(1), 71-80.

Brath, A., Castellarin, A., Franchin, M. and Galeati, G. (2001) Estimating the index flood using indirect method, Hydrological Sciences Journal, 46(3), 399-418.

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. and Goel, N. K. (2000) The formation of group for regional flood frequency analysis, Hydrological Sciences Journal, 45(1), 97-112.

Bullock, A., Andrews, A. and Mngodo, R. (1997) Regional surface water resources and drought assessment, Southern African FRIEND. Technical Documents in Hydrology No 15, UNESCO, Paris.

Campolo, M., Andreeussi, P. and Soldati, A. (1999) River flood forecasting with a neural network model. Water Resources Research, 35(4), 1191-1197.

Central Water Commission of India, (1989) Manual on Flood Forecasting. River Management Wing, New Delhi

Chang, Fi-John, Chiang, Yen-Ming and Chang, Li-Chiu, (2007) Multi-step ahead neural networks for flood forecasting. Hydrological Sciences Journal, 52(1), 114-130.

Chang, F.J. and Chen, Y.C. (2001) A counterpropagation fuzzy neural network modeling approach to real time streamflow prediction. Journal of Hydrology, 245, 153-164.

Chau, K.W. (2006) Particle swarm optimization training algorithm for ANNs in stage prediction in Shing Mun river, Journal of Hydrology, 329(3-4), 363-367.

Chavoshi,S. and Soleiman, W.N.A. (2009) Delineating pooling group for flood frequency analysis using soft computing. European Journal of Scientific Research, ISSN 1450-216X, 35(2), 181-187.

Chen, S., Cowan, C.F.N. and Grant, P.M. (1991) Orthogonal least squares learning algorithm for radial basis function networks. IEEE Transactions on Neural Networks 2(2), 302–309.

Chen, Y.H.,Wang, W.J. and Chiu, C.H. (2000) New estimation method for the membership values in fuzzy set. Fuzzy Sets and Systems, 112, 521-525.

Cheng, C.T., Ou, C.P. and Chau, K.W. (2002) Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration, 268(1-4), 72-86.

Chiu, S.L.(1994) Fuzzy model identification based on cluster estimation. Journal on Intelligent Fuzzy Systems, 2, 267-278.

Chowdhary, A.K. (2005) Regional flood frequency analysis: A case study of Mahanadi basin, Proceeding of GLOGIFT 05, RGPV, Bhopal, 456-461.

Coulibaly, P., Anctil, F. and Bobe´e, B. (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach, Journal of Hydrology, 230(3-4), 244-257.

Dawson, C.W. and Wilby, R.L. (1998) An artificial neural network approach in rainfall-runoff modeling. Hydrological Sciences Journal, 43(1), 47-66.

Dawson, C.W. and Wilby, R.L. (2001) Hydrological modeling using Artificial Neural Networks. Progress in Physical Geography, 25(1), 80-108.

Dawson, C.W., Abrahart, R.J., Shamseldin, A.Y. and Wilby, R.L. (2006) Flood estimation at ungauged sites using artificial neural network. Journal of Hydrology, 319(1-4), 391-409.

DeCoursey, D. G. (1973) Objective regionalization of peak flow rates. In flood and Droughts. E. F. Koelzer and K. Mahmood (Editors), Proceedings of the second International Symposium in Hydrology, September 11-13, 1972, Fort Collins, Colorado. Water Resources Publications, 395-405.

Decoursey, D. G. and Deal, R.B. (1974) General aspects of multivariate Analysis with Application to some problems in Hydrology. Proceedings of Symposium on Statistical Hydrology. Misc.Pub.No.1275, USDA-ARS, 47-68.

Deshmukh, R.P. and Ghatol, A.A. (2010) Short term flood forecasting using RBF static neural network modeling a comparative study. International Journal of Computer Science and Information Security, 8(6), 93-98.

Dharmasena, G.T. (1997) Application of mathematical models for flood forecasting in Srilanka, Destructive Water: Water-caused natural disasters, their abatement and control (Proceeding of the conference held at Anaheim, California, June, 1996), IAHS, Publ.No.239, 225-235.

Dikbas, F., Firat, M., Cem Koc, A. and Ghungor, M. (2011) Classification of precipitation series using fuzzy cluster method. International Journal of Climatology, DOI: 10.1002/joc. 2350.

Dou, C., Woldt, W. and Bogardi, I. (1999) Fuzzy rule based approach to describe solute transport in the unsaturated zone. Journal of Hydrology, 220(1-2), 74–85.

Dubois, D., Nguyen, H.T., Prade,H. and Sugeno, M. (1998) Introduction: the real contribution of fuzzy systems, in: Nguyen, H.T and Sugeno, M.(Eds.) Fuzzy systems modeling and control, Kluwer, Dordrecht, 1-17.

Eslamian, S.S. and Feizi, H. (2007) Maximum monthly rainfall analysis using L-moments for an arid region in Isfahan province, Iran. Journal of Applied Meteorology and Climatology, 46,494-503.

Eslamian, S.S. and Biabanaki, M. (Fall 2008) Low flow regionalization modeling. International Journal of Ecological Economics and Statistics, 12(F08), 82-97.

Fernando, A.K., Jayawardena, A.W. (1998) Runoff forecasting using RBF networks with OLS algorithm. Journal of Hydrologic Engineering, 3(3), 203–209.

Firat, M. and Gungor, M. (2008) Hydrological time series modeling using an adaptive neuro-fuzzy inference system. Hydrological Processes, 22(13), 2122-2132.

Flood, I. and Kartam, N. (1994) Neural networks in civil engineering, I: Principles and understandings. Journal of Computing in Civil Engineering, 8(2), 131-148.

Freeze, R. A. and Cherry, J. A. (1979) Groundwater hydrology. Prentice-Hall Inc., Englewood Cliffs, New Jersy, 604p.

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.

Gogoi, S. and Joshi, R.C. (2011) Flood forecasting with the help of fuzzy rule based modeling of Jiadhal river basin, Dhemaji, Assam. Advances in Fuzzy Mathematics, 6(1), 155-164.

Gupta, P.N. (1980) Regional flood frequency analysis approach-estimation of peak floods for Mahanadi basin (sub zone 3d). Indian Railway Technical Bulletin, February, 29-39.

Haghizadeh, A., Teang shui, L. and Goudarzi, E. (2010) Estimation of Yield Sediment Using Artificial Neural Network at Basin Scale. Australian Journal of Basic and Applied Sciences, 4(7), 1668-1675.

Halff, A.H., Halff, H.M. and Azmoodeh, M. (1993) Predicting runoff from rainfall using neural networks. In engineering Hydrology, Kuo CY (ed.) Proceeding of the Symposium sponsored by the Hydraulics Division of ASCE, San Fransisco, CA, July 25-30,1993, ASCE: New York, 760-765.

Hall, A.J. (1972) Methods of selection of areal rainfall stations and the calculation of areal rainfall for flood forecasting purposes. Melbourne: Commonwealth Bureau of Meteorology, Melbourne, 27, [10]p.

Hall, M.J and Minns, A.W. (1999) The classification of hydrologically homogeneous regions. Hydrological Sciences Journal, 44(5), 693-704.

Hamerly, G. and Elkan, C., (2002) Alternatives to the k-means algorithm that find better clusterings. In: Kalpakis, K., Goharian, N., Grossmann, D. (Eds.), Proceedings of the Eleventh International Conference on Information and Knowledge Management, (CIKM-02). ACM Press, New York, 600–607.

Hellendoorm, H. and Driankov, D. (Eds.) (1997) Fuzzy model identification: selected approaches. Springer, Berlin, Germany.

Hong, Y-S.T., and White, P.A., (2009) Hydrological modeling using a dynamic neuro-fuzzy system with on-line and local learning algorithm, Advances in Water Resources, 32(1), 110–119, 2009.

Hosking, J.R.M. (1986) L-moments: analysis and estimation of the theory of probability weighted moments. Res. Rep. RC 12210, IBM Research Division, Yorktown Heights, N.Y.

Hosking, J.R.M. (1990) L-moments: Analysis and estimation of distribution using linear combinations of order statistics. J. R. Stat. Soc., SER. B., 52(1), 105-124.

Hosking, J..R.M. and Wallis J.R. (1991) Some Statistics useful in Regional frequency analysis. Res. Rep. RC 17096, IBM Research Division, Yorktown Heights, N.Y.

Hosking, J.R.M. and Wallis, J.R. (1993) Some Statistics useful in Regional frequency analysis. Water Resources Research, 29(2), 271-282.

Hosking, J. R. M. and Wallis J. R. (1997) Regional Frequency Analysis: An approach based on L-moments. Cambridge University Press, Cambridge.

Hsu, K.L., Gupta, H.V. and Sorooshian, S. (1995) Artificial neural network modeling of the rainfall-runoff process. Water Resources Research, 31(10), 2517-2530.

Hsu,K.L., Gupta,H.V., Gao,X.,Sorooshian,S. and Imam,B. (2002) Self organisationlinear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, Water Resource Research, 38(12), 38-1.

Hundecha, Y., Bardossy, A. and Theisen, H. (2001) Development of a fuzzy logic-based rainfall-runoff model. Hydrological Sciences Journal, 46(3), 363-376.

Imrie, C.E., Durucan, S. and Korre, A. (2000) River flow prediction using artificial neural network: Generalization beyond the calibration range. Journal of Hydrology, 233(1-4), 138-153.

Jacquin, A. P. and Shamseldin, A.Y. (2006) Development of rainfall-runoff models using Takagi-Sugeno fuzzy inference systems, Journal of Hydrology, 329 (1-2), 154-173.

Jain, A. and Indurthy, S.K.V.P. (2003) Comparative analysis of event based rainfall-runoff modeling techniques, deterministic, statistical and artificial neural networks. Journal of Hydrologic Engineering, 8(2), 93-98.

Jang, J-S.R.(1993) ANFIS: Adaptive network based fuzzy inference system. IEEE Transactions on Systems, Man and Cybernetics, 23(3), 665-685.

Jang, J-S.R. and Mizutani, E. (1996) Levenberg-Marquardt method for ANFIS learning. Proceeding of the International Joint Conference of the North American Fuzzy Information Processing Society Bi-annual Conference, Berkely, California 87-91.

Jayawardena, A.W., Fernando, D. A. K., and Zhou, M. C. (1996) Comparison of multi-layer perceptron and Radial Basis Function Networks as tools for flood forecasting, Destructive water: Water-Caused Natural Disasters, their Abatement and Control (Proceedings of the Conference held at Anaheim, California, June 1996). IAHS Publ. No. 239, 173-181.

Jingyi, Z. and Hall, M.J. (2004) Regional flood frequency analysis for the Gan-Ming river basin in China. Journal of Hydrology, 296(1-4), 98–117.

Kar, A.K., Lohani, A.K., Goel, N.K. and Roy, G.P. (2009) Flood problem of Mahanadi basin: Issues and measures. National symposium on Climate Change and Water Resources in India (CCWRIN) organized by Indian Association of Hydrologists, Nov.18-19, 2009, Roorkee.

Kar, A.K., Lohani, A.K., Goel, N.K. and Roy, G.P. (2010) Development of flood forecasting system using statistical and ANN techniques in the downstream catchment of Mahanadi basin. Journal of Water Resources Protection, 2(10), 880-887.

Kagan, R.L. (1966) On the evaluation of representativeness of RG data. G1. Geofiz. Obs. 191, 22-34.

Kaiser, H. F. (1960) The application of electronic computers to factor analysis. Educ. Psychol. Meas., 20,141-151.

Keskin, F. (2008) Testing of different meteorological models for flood forecasting in Filyos basin. Turkey, BALWOIS, Ohrid, Republic of Macedonia, May27-31, 2008.

Kevin, D.Y., Kibler, D.F., Benham, B.L. and Loganathan, G.V. (2009) Application of analytical hierarchical process for improved selection of storm water BMPs. Journal of Water Resources Planning and Management, 135(4), 264-275.

Khatua, K.K. and Mahakul, B. (1999) Flood in Mahanadi delta stage-II area- a case study. Proceeding of National Seminar on Disaster Management, UCE, Burla, Orissa, 12-13 Nov.1999.

Kiang, M.Y., Kulkarni, U.R., Goul, M.R., Philppakkis, A., Chi, R.T. and Turban, E. (1997) Improving the effectiveness of self organization map networks using a circular Kohonen layer, 30th Hawaii International Conference on System Sciences (HICSS) Volume 5, 521-529, Advanced Technology Track, 1997.

Kisi, O. (2004) River flow modeling using artificial neural networks. Journal of Hydrologic Engineering, 9(1), 60-63.

Kisi, O. (2007) Streamflow forecasting using different artificial neural network algorithms, Journal of Hydrologic Engineering, 12(5), 532-539.

Kisi, O. and Cigizoglu, H.K. (2007) Comparison of different ANN techniques in continuous and intermittent river flow prediction, Civil Engineering and Environment Systems, 24(3), 211-231.

Kisi, O. (2008) River flow forecasting and estimation using different artificial neural network techniques. Hydrology Research, 39(1), 27-40.

Kisi, O. (2009) Neural networks and wavelet conjunction model for intermittent streamflow forecasting. Journal of Hydrologic Engineering, 14(8), 773-782.

Kisi, O. (2010) Fuzzy genetic approach for modeling reference evapotranspiration, Journal of Irrigation and Drainage Engineering, 136(3), 175-183.

Kohonen, T. (1997) Self Organisation Map (2nd Edition). Springer, Berlin ISBN3-540-62017-6.

Kruse, R., Gebhardt, J. and Klawonn, F. (1994) Foundations of Fuzzy Systems. Wiley, Chichester.

Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W. and Pruitt, W. O, (2002) Estimating evapo-transpiration using artificial neural network. Journal of Irrigation and Drainage Engineering, 128(4), 224-233.

Kumar, R., Singh, R. D. and Seth, S. M. (1999) Regional flood formula for seven subzones of India. Journal of Hydrologic Engineering, 4(3), 240-244.

Kumar, R., Chatterjee, C., Kumar, S., Lohani, A. K. and Singh, R. D. (2003) Development of Regional Flood Frequency Relationships using L-moments for Middle Ganga Plains Subzone 1(f) of India. Water Resources Management, 17, 243–257.

Kumar,R. and Chatterjee, C. (2005) Regional flood frequency analysis using L-moment for North Brahmaputra region of India. Journal of Hydrologic Engineering, 10(1), 1-7.

Kumar,R. (2009) Regional flood frequency estimation in India, Ph.D. Thesis, IIT,Roorkee.

Lekkas, D.F., Onof, C., Lee, M.J. and Baltas, E.A. (2004) Application of Artificial neural networks for flood forecasting. Global Nest: The Int. Journal. 6 (3), 205-211.

Lim, Y.H. and Lye, L.M. (2003) Regional Flood Estimation for Ungauged Basins in Sarawak, Malaysia, Hydrological Sciences Journal, 48(1), 79-94.

Liong, S.Y., Lim, W.H., and Paudyal, G. N. (2000) River stage forecasting in Bangladesh: Neural network approach. Journal of Computing in Civil Engineering, 14(1), 1–8.

Lohani, A.K., Goel, N.K. and Bhatia, K.K.S. (2005a) Real time flood forecasting using fuzzy logic. Hydrological Perspectives for Sustainable Development. Vol. 1, Perumal M (ed). Allied Publishers Pvt. Ltd.: New Delhi, 168–176.

Lohani, A.K., Goel, N.K., Bhatia, K.K.S. (2005b) Development of fuzzy logic based real time flood forecasting system for river Narmada in Central India, In International Conference on Innovation Advances and Implementation of Flood Forecasting Technology. ACTIF/ Floodman / Flood Relief, October, 2005. Tromso,Norway; www. Actif. cc. net/ conference 2005 / proceedings.

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.

Lohani, A.K., Goel, N.K. and Bhatia, K.K.S. (2007a) Deriving stage-discharge-sediment concentration relationships using fuzzy logic. Journal of Hydrological Sciences, 52(4): 793–807.

Lohani, A.K., Goel, N.K.and Bhatia, K.K.S. (2007b) Reply to comments provided by Z. Sen on Takagi–Sugeno fuzzy system for modeling stage discharge relationship by A.K. Lohani, N.K. Goel and K.K.S. Bhatia. Journal of Hydrology, 337(1–2): 244–247.

Lohani, A.K., Goel, N.K. and Bhatia, K.K.S. (2011) Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modeling under different input domains. Hydrological Processes, 25,175-193.

Luchetta, A. and Manetti, S. (2003) A real time hydrological forecasting system using a fuzzy clustering approach. Computers and Geosciences, 29(9), 1111-1117.

MacQueen, J. B. (1967) Some methods for classification of multivariate observation. Proc.5th Berkeley Symposium on Probability and Statistics. University of California Press, Berkely, 281-297.

Machiwal, D., Jha, M.K. and Mal, B.C. (2011) "Assessment of ground water potential, in a semi-arid region of India using Remote Sensing, GIS and MCDM Techniques. Water Resources Management, 25(5), 1359-1386.

Maier, H.R. and Dandy, G.C. (2000) neural networks for prediction and forecasting of water resources variables: A review of modeling issues and applications. Environmental Modelling and Software, 15(2000), 101-124.

Malekinezhad, H., Nachtnebel, H.P. and Klik, A. (2011) Comparing the index flood and multiple regression method using L-moment. Physics and Chemistry of the Earth, 36(1-4), 54-60.

Mamdani,E.H. and Assilian,S. (1975) An experiment with linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies, 7(1),1-13.

Mason, J.C., Price, R.K. and Tem'me, A. (1996) A neural network model of rainfall-runoff using radial basis functions. Journal of Hydraulic Research, 34(4), 537–548.

Math Works(1994) The Math Works MATLAB Digest. 2(5).

Mayilvaganan, M.K. and Naidu, K.B. (2011) ANN and Fuzzy logic models for the prediction of ground water level of a watershed. IJCSE, 3(6), 2523-2530.

McCulloch, W .and W. Pitts., W. (1943) A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5(4), 115 – 133.

Mehrotra, K. Mohan, C.K. and Ranka, S. (1997) Elements of artificial neural network, The MIT Press, Massachussets Institute of Technology, Massachussets 02142.

Minns, A. W. and Hall, M. J. (1996) Artificial neural networks as rainfall-runoff models. Hydrological Sciences Journal, 41(3), 399-417.

Moody, J. and Darken, C. (1989) Fast learning in networks of locally tuned processing units. Neural Computation 1, 281–294.

Morin, G., Fortin, J. P., Sochanska, W. and Lardeau, J. P. (1979) Use of Principal component analysis to identify homogeneous stations for optimal interpolation. Water Resources Research, 15(6), 1841-1850.

Mukerji, A., Chatterjee, C. and Raghuwanshi, N. S. (2009) Flood forecasting using ANN, Neuro-Fuzzy and Neuro-GA models. Journal of Hydrologic Engineering, 14(6), 647-652.

Nayak, P.C., Sudheer, K.P., Rangan, D.M. and Ramashastri, K.S. (2004) A neuro-fuzzy computing technique for modeling hydrological time series. Journal of Hydrology, 291, 52-66.

Nayak, P.C., Sudheer, K.P. and Ramashastri, K.S. (2005) Fuzzy computing based rainfall-runoff model for real time flood forecasting. Hydrological Process.19, 955-968.

Nayak, P.C., Sudheer, K.P. and Jain, S.K. (2007) Rainfall–runoff modeling through hybrid intelligent system, Water Resources Research **43**.

Nayak, P.C. (2010) Explaining internal behavior in a fuzzy if–then rule-based flood-forecasting model. Journal of Hydrologic Engineering, 15(1), 20-28.

Nandalal, H.K. and Ratanayake, U.R. (2011) Flood risk analysis using fuzzy models. Journal of Flood Risk Management, 4(2), 128-139.

Nathan, R. J. and McMahon, T. A. (1990) Identification of homogeneous regions for the purpose of regionalization. Journal of Hydrology, 121(1-4), 217–238.

National Institute of Hydrology (1997-98) Regional flood frequency analysis using L-moment. Technical Report, TR (AR), NIH, Roorkee.

National Water Development Agency (2004) Technical Study No.174. Water balance study of Mahanadi basin upto Hirakud dam (Revised).

Nauck, D. and Kruse, R. (1997) A neuro-fuzzy method to learn fuzzy classification rules from data. Fuzzy Sets and Systems, 89, 277-288.

NERC, (1975) Flood Studies Report, Environ. Res. Council., London, 1-5, 1100.

Ni, J.R. and Xue, A. (2003) Application of artificial neural network to the rapid feedback of potential ecological risk in flood diversion zone. Engineering Applications of Artificial Intelligence, 16(2), 105-119.

Nielsen, S.A. and Hanseen, E. (1973) Numerical simulation of rainfall-runoff process on a daily basis, Nordic Hydrology, 4, 171-190.

Nor, N.A., Harun, S. and Kasim, A.H. (2007) Radial basis function modeling of hourly streamflow hydrograph, journal of Hydrologic Engineering,12(1), 113-123.

Ouarda, T.B.M.J., & C. Shu (2009) Regional low-flow frequency analysis using single and ensemble artificial neural networks, Water Resour. Res., 45, W11428, doi:10.1029/2008WR007196.

Ozelkan, E.C. and Duckstein, L. (2001) Fuzzy conceptual rainfall-runoff models. Journal of Hydrology, 253(1-4), 41-68.

Pang, B., Guo, S.L., Xiong, L.H. and Li, C.Q. (2007) A nonlinear perturbation model based on artificial neural network, Journal of Hydrology, 333(2-4), 504–516.

Parida, B. P. (2004) A partitioning methodology for identification of homogeneous regions in regional flood frequency analysis. BALWOIS 2004, Orchid, FY Republic of Macedonia, 25-29 May, 2004.

Parida, B. P., Moalafhi, D. B. and Kenabatho, P. K. (2006) Forecasting runoff coefficients using ANN for water resources management: The case of Notwane catchment in Eastern Botswana. Physics and Chemistry of the Earth, 31, (15-16), 928-934.

Park, J. and Sandberg, I. W. (1991) Universal approximations using Radial-Basis-Function networks. Neural Computation, 3(2), 246-257.

Park, J.H. (2010) A study on the comparative method using AHP and GIS based distributed runoff model. KSCE Journal of Civil Engineering, 14(6), 953-960.

Patra, K. C. (2002) Hydrology and water resources engineering. Narosa Publishing House, New Delhi.

Pedrycz, W. (1984) An identification algorithm in fuzzy relational systems. Fuzzy Sets and Systems, 13, 153-167.

Pillai, C.R.S. and Raju, K.S. (1996) Ranking irrigation management alternatives by multicriterion analysis. Water Resources Development, 12(3), 329-345.

Pitlick, J. (1994) Relation between peak flows, precipitation, and physiography for five mountainous regions in the western USA. Journal of Hydrology, 158, 219-240.

Rabuffetti, D. and Barbero, S. (2005) Operational hydro-meteorological working and real time flood forecasting: the Piemonte region case study. Hydrology and Earth System Sciences, 9(4), 457-466.

Raghuwanshi, N. S, Singh, R. and Reddy, L.S. (2006) Runoff and sediment yield modeling using artificial neural networks: Upper Siwane River, India, Journal of Hydrologic Engineering, 11 (1), 71–79.

Rajsekharan, S. and Vijaylakshmi Pai, G.A. (2004) Image recognition using simplified fuzzy artmap augmented with a moment based feature extractor. Pattern Recognition, artificial Intelligence, 14(8), 1081-1094.

Raju, K.S., Duckstein, L. and Arondel, C. (2000) Multi-criterion analysis for sustainable water resources planning: a case study in Spain. Water Resources Management, 14:435-456.

Raju, K.S. and Kumar, D.N. (1999) Multicriterion decision making in irrigation planning. Agricultural Systems, 62(2):117-129.

Raju, K.S. and Pillai, C.R. (1999) Multicriterion decision making in river basin planning and development. European Journal of Operational Research, 112(2):249-257.

Raju, K.S. and Pillai, C.R.S. (1999) Multicriterion decision making in performance evaluation of an irrigation system. European Journal of Operational Research, 112(3), 479-488.

Raju, K.S. and Pillai, C.R.S. (1999) Multicriterion decision making in river basin planning and development. European Journal of Operational Research, 112(2), 249-257.

Raju, K.S., nagesh Kumar, D. and Duckstein, L. (2006) Artificial neural network and multi-criterion analysis for sustainable irrigation planning. Computers and Operation Research, 33, 1138-1153.

Rajurkar, M.P., Kothyari, U.C. and Chaube, U.C. (2002) Artificial neural networks for daily rainfall-runoff modeling, Journal of Hydrological Sciences, 47(6), 865-877.

Rao A. R. and Hamed, H.K. (2000) Flood Frequency Analysis. CRC Press, Boca Raton Florida, U.S.

Rao, A.R. and Srinivas, V.V. (2006) Regionalisation of watersheds by fuzzy cluster analysis. Journal of Hydrology, 318(1-4), 57-79.

Robson, A. and Reed, D. (1999) Statistical procedure for flood frequency estimation. Flood Estimation Handbook.

Saf, B. (2008) Application of index procedures to flood frequency analysis in Turkey. Journal of the American Water Resources Association, 44(1), 37–47.

Saf, B. (2009) Regional flood frequency analysis using L-moments for the Buyuk and Kucuk Menderes River Basins of Turkey. Journal of Hydrologic Engineering, 14(8), 783–794.

Saaty, T.L. (1980) Analytic Hierarchy Process, McGraw Hill, New York, 20-25.

Saaty, T.L. (2008) Decision making with analytic hierarchy process. International Journal of Services Sciences, 1(1), 83-98.

Sahoo, G.B. and Ray, C. (2006) Flow forecasting for a Hawaii system using rating curves and neural networks. Journal of Hydrology, 317(1-2), 63-80.

Sarkar, S., Goel, N. K. and Mathur, B.S. (2010) Development of iso-pluvial map using L-moment approach for Tehri - Garhwal Himalaya. Stoch. Environ. Res. Risk. Assess, 24(3), 411-423.

See, L. and Openshaw, S. (1999) Applying soft computing approaches to river level forecasting, Hydrological Sciences Journal, 44(5), 763-778.

Senthil Kumar, A. R., Sudheer, K.P., Jain, S.K. and Agarwal, P.K. (2005) Rainfall-runoff modeling using artificial neural networks: Comparison of network types. Hydrological Processes, 19(6), 1277-1291.

Shi, J.J. (2002) Clustering techniques for evaluating and validating neural network performance, Journal of Computing in Civil Engineering, 16(2), 152-155.

Shi,Y and Mizumoto, M. (2001) An improvement of neuro-fuzzy learning algorithm for tuning fuzzy rules. Fuzzy Sets and Systems, 118, 339-350.

Shibata R. (1976) Selection of the order of an autoregressive model by Akaike's information criteria. Biometrika 63, 117–126.

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, 132–149.

Shu, C. and Ouarda, T.B.M.J. (2008) Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. Journal of Hydrology, 349, 31– 43.

Shumway, R.H. (1988) Applied Statistical Time Series Modeling. Prentice Hall: Englewood Cliffs, NJ.

Singh, R.D. and Seth, S.M. (1985) Regional flood frequency analysis for Mahanadi basin using wakeby distribution. Proceedings of seminar on Flood Frequency Analysis, New Delhi.

Sinha, R., Bapalu, G.V., Singh, L.K. and Rath, B. (2008) Flood risk analysis in the Kosi river basin, north Bihar using multi-parametric approach of Analytical Hierarchy Process (AHP), Indian Journal of Remote Sensing, 36, 293-307.

Sreedharan, K.E. and James, E.J. (1983) Design of rain gage network for the Chaliyar River basin. Journal of Institute of Engineers. (India).64 (CI 3), 170-174.

Stambuk, A., Stambuk, N. and Konjevoda, P. (2007) Application of Kohonen Self-Organising Maps (SOM) based clustering for the Assessment of Religious Motivation. Information Technology Interfaces, 87-91.

State Water Plan (2007). DOWR. Govt. of Orissa.

Stedinger, J.R., Vogel, R.M. and Foufoula, G.E. (1992) Frequency analysis of extreme events. Handbook of Hydrology, Maidment, Chapter 18, Mc Graw Hill, N.Y.

Sudheer, K.P. (2000) Modeling hydrological processes using neural computing technique. PhD thesis, Indian Institute of Technology, Delhi, India.

Sudheer, K.P., Gosain, A.K. and Ramashastri, K.S. (2002, a) A data driven algorithm for constructing artificial neural networks rainfall-runoff models. Hydrological processes, 16(6), 1325-1330.

Sudheer, K.P., Gosain, A.K. and Ramasastri, K.S. (2002, b) Comparisons between back propagation and radial basis function based neural networks in rainfall-runoff modeling. In Proceedings of the International Conference on Advances in Civil Engineering, Vol. 1, Kharagpur, India, 449–456.

Takagi, T. and Sugeno, M. (1985) Fuzzy identification of systems and its application to modeling and control. Transctions on Systems, Man and Cybernetics, IEEE, 15(1), 116-132.

Tareghian, R. and Kashefipour, S.M. (2007) Application of fuzzy systems and artificial neural networks for flood forecasting. Journal of Applied Sciences, 22(7), 3451-3459.

Tasker, G.D. (1982) Comparing methods of hydrologic regionalization. Water Resources Bulletin, 18(6), 965-970.

Teschl, R. and Randeu, W. L. (2006) A neural network model for short term river flow prediction. Natural Hazards Earth System and Sciences, 6, 629-635.

Thandeveswara, B. S. and Sajjikumar, N. (2000) Discussion on Classification of River Basins using Artificial Neural Network. Journal of Hydrologic Engineering, 5(3), 290 -298.

Theodoridis, S. and Koutroumbas, K. (1999) Pattern Recognition. Academic Press, New York; 482–483.

Tingsanchali,T. and Gautam, M.R. (2000) Application of tank, NAM, ARMA and neural network models to flood forecasting. Hydrological Processes, 14(10-14), 2473-2487.

Thirumalaiah, K. and Deo, M. C. (1998) Real-Time flood forecasting using neural networks. Computer-Aided Civil and Infrastructure Engineering. 13, 101–111. doi: 10.1111/0885-9507.00090

Tiwari, M.K. and Chatterjee, C. (2010) Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. Journal of Hydrology, 394(3-4), 458-470.

Tokar, A.S. and Johnson, P.A. (1999) Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering, 4(3), 232–239.

Triantaphyllou, E. and Mann, S.H. (1995) Using the analytic hierarchy process for decision making in engineering applications: Some challenges. International Journal of Industrial Engineering: Applications and Practice, 2(1), 35–44.

Tsiko, R. G. and Haile, T. S. (2011) Integrating Geographic Information System, Fuzzy logic and Analytical hierarchical process in modeling optimum sites for locating water reservoirs. A case study of Debub district of Eritrea. Water, 3, 254-290.

Tsukamato, Y. (1979) An approach to fuzzy reasoning method. In Advances in Fuzzy Set Theory and Application, edited by M. M. Gupta, R. K. Ragade and R. R. Yager, 137-149, North- Holland, New York.

Turan, M. E. and Yurdusev, M. A. (2008) Unmeasured hydrologic data estimation by artificial intelligence methods. BALWOIS 2008, Ohrid, Republic of Macedonia, 27-31 May 2008.

Uvo, C.B., Tölle, U. and Berndtsson, R. (2000) Forecasting discharge in Amazonia using artificial neural networks. International Journal of Climatology, 20(12), 1495 – 1507.

Vernieuwe, H., Georgieva, O., Baets, B. De., Pauwels, V.R.N., Verhoest, N.E.C. and Torch, F. P. De. (2005) Comparison of data driven Takagi-Sugeno models of rainfall discharge dynamics. Journal of Hydrology, 302(1-4), 173-186.

Vogel, R.M. and Kroll, C.N. (1992) Regional geohydrologic-geomorphic relationships for estimation of low flow statistics. Water Resources Research, 28(9), 2451-2458.

Vogel, R.M, Wilson, I. and Daly, C. (1999) Regional regression models of annual streamflow for the United States. Journal of Irrigation and Drainage Engineering, 125(3), 148-157.

Xiong, L., O'Connor, K.M. and Guo, S. (2004) Comparison of the three updating schemes using ANN in flow forecasting. Hydrology and Earth System Sciences, 8(2), 247-255.

World Meteorological Organization, WMO (1994) Guide for hydrological practices. Geneva, Switzerland.

Wu, J.S., Han, P.E., Annambhotla, S. and Bryant, S. (2005) Artificial neural networks for forecasting watershed runoff and stream flows. Journal of Hydrologic Engineering, 10(3), 216-222.

Wu, C.L. and Chau, K.W. (2006) Evaluation of several algorithms in forecasting flood. Lecture notes in Computer Science, 4031 LNCS, 111-116.

Wu, C.L., Chau, K.W. and Fan, C. (2010), Prediction of rainfall time series using modular artificial neural network coupled with data processing techniques, Journal of Hydrology, 389(1-2), 146-167.

Zadeh, L. A. (1965) Fuzzy Sets, Information and control. 8(3), 338-353.

Zadeh, L. A. (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man and Cybernetics,1,28-44.

Zadeh, L. A. (1998) Some reflections on the anniversary of fuzzy sets and systems, Fuzzy Sets and Systems, 100, 5-7.

Zealand, C.M., Burn, D.H. and Simonovic, S.P. (1999) Short-term stream flow forecasting using artificial neural network. Journal of Hydrology, 214(1-4), 32-48.

ANNEXURE- I

Losses incurred during floods at Orissa (1960-2008)

Sl. No.	Year	Month of occurrence	Rivers	Affected District / Area	Loss/Damage Reported		
					Human	Live Stock	Public Utility
1	1960	Aug.	Mahanadi, Brahmnai, Baitarani, Burhabalanga & Subarnarekha	Cuttack, Puri, Dhenkanal, Balasore, Mayurbhanj & Keonjhar - **6districts.**	Not Available	Not Available	18.65 lakh Ac. of cropped area damaged and loss of Rs.11.70 crores.
2	1961	July-Sep	Mahanadi, Brahmnai, Baitarani, Burhabalanga & Subarnarekha	Cuttack, Puri, Dhenkanal, Balasore, Mayurbhanj & Keonjhar - **6districts.**	Not Available	Not Available	1.429 lakhs Ac. of cropped area damaged with a loss of Rs.2.54 crores.
3	1964	July-Aug	Mahanadi, Brahmani, Baitarani & Rushikulya	Cuttack, Puri, Bolangir, Dhenkanal, Balasore, Sambalpur, Ganjam, Phulbani & Keonjhar - **9districts.**	Not Available	Not Available	4.08 lakh Ac. of cropped area damaged.
4	1971	July-Oct	Mahanadi, Brahmnai, Baitarani & Subarnarekha	Cuttack, Balasore, Puri,Mayurbhanj, Bolangir, Sundergarh& Keonjhar - **7 districts.**	26	265	11.719 lakh Ac. of cropped area damaged. 95043 no. of houses damaged. Total loss of Rs.31.71 crores.
5	1974	Aug	Mahanadi, Brahmnai, Baitarani, Burhabalanga & Subarnarekha	Cuttack, Balasore, Puri, Dhenkanal & Keonjhar - **5districts.**	Not Available	Not Available	5.40 lakh Ha. of cropped area damaged.
6	1980	Sept	Mahanadi, Brahmnai, Baitarani & Vamsadhara	Balasore, Bolangir, Cuttack, Dhenkanal, Ganjam, Kalahandi, Koraput, Phulbani, Puri, Sambalpur **- 10 districts.**	82	16669	3.19 lakh Ha. of cropped area damaged. Rs.65.00 crores of PU damaged.
7	1982	Aug Sept.	Mahanadi, Rushikulya	Cuttack, Puri, Bolangir, Phulbani, Ganjam, Sambalpur, Dhenkanal & Kalahandi - **8 districts.**	126	26359	12.00 lakh Ha. of cropped area damaged. Rs.616.00 crore of PU damaged.
8	1984	Jun-Sept.	Subarnarekha, Brahmani, Baitarani, Mahanadi, Vamsadhara, Indrabati	Cuttack , Balasore, Puri, Phulbani, Koraput, Ganjam , Dhenkanal, Keonjhar - **8 districts.**	27		3.92 lakh Ha. of cropped area damaged.
9	1985	Aug - Sept.	Mahanadi, Rushikulya, Baitarani, Brahmani, Subarnarekha,	Balasore, Bolangir, Cuttack, Ganjam, Puri, Phulbaani, Keonjhar, Kalahandi, Sambalpur - **9districts**	22	5281	3.10 lakh Ha. of cropped area damaged.
10	1986		Mahanadi, Subarnarekha, Indravati	Balasore, Bolangir, Cuttack, Dhenkanal, Koraput, Mayurbhanj, Puri, Phulbaani, Sambalpur- **9 districts.**	24	337	1.08 lakh Ha. of croppedarea damaged. Rs.55.31 crore of PU damaged.
11	1991	July-Aug.	Mahanadi, Brahmani, Baitarani, Subarnarekha, Vamsadhara	Cuttack, Puri, Balasore, Ganjam, Phulbani, Dhenkanal, Sambalpur, Kalahandi, Koraput, Keonjhar- **10 districts.**	52	1145	6.62 lakh Ha. of cropped area damaged.
12	1992	Jun-Aug.	Mahanadi,	Cuttack, Puri, Bolangir, Balasore,	43	1397	4.17 lakh Ha. of

			Subarnarekha, Vamsadhara	Ganjam, Koraput, Phulbani, Samabalpur, Kalahandi, Dhenkanal, Sundergarh- **11 districts**			cropped area damaged. Rs.184.48 crore of PU damaged.
13.	1994	July-Sept.	Mahanadi, Brahmani, Subarnarekha, Vamsadhara	Angul, Balasore, Bhadrak, Boudh, Cuttack, Jagatsinghpur, Jajpur, Jharsuguda, Khurda, Kendrapara, Kalahandi,Koraput, Malkanagiri, Nayagarh, Nowrangpur, Puri, Sambalpur, Sundergarh, Sonepur- **20 districts.**	50		10.17 lakh Ha. of cropped damaged.
14	1995	May-Nov.	Mahanadi, Subarnarekha, Vamsadhara, Rushikulya	Angul, Balasore, Bhadrak, Boudh, Cuttack, Dhenkanal, Ganjam, Gajapati, Jagatsinghpur, Jajpur, Khurda, Koraput Kalahandi,Kendrapara, Keonjhar, Kandhamal,Malkangiri, Nawarangpur, Nayagarh, Puri, Rayagada, Sambalpur, Sonepur **- 23districts.**	76	372	16.09 lakh Ha. of cropped area damaged. Rs.112.42 crore of PU damaged.
15	1997	Jun-Aug.	Mahanadi	Balasore, Bhadrak, Cuttack, Denkanal, Jagatsinghpur, Jajpur, Khurda, Kalahandi, Kendrapara, Kandhamal, Keonjhar, Mayurbhanj, Nuapara, Nawarangpur, Nayagarh, Puri, Sundergarh, Sambalpur- **18 districts.**	29	52	5.27 lakh Ha. of cropped area damaged.
16	1999	July-Aug.	Mahanadi, Brahmani, Baitarani, Subarnarekha, Rushikulya	Cuttack, Jagatsinghpur, Kendrapara, Jajpur, Bhadrak, Balasore, Mayurbhanj- **7 districts.**	10		1.49 lakh Ha. of cropped area damaged. Rs.54.00 crores of PU damaged.
17	2001	July-Aug.	Mahanadi, Brahmani, Baitarani, Subarnarekha, Burhabalanga, Vamsadhara, Rushikulya, Indravati	Angul, Balasore, Bhadrak, Boudh, Bolangir, Baragarh, Cuttack, Dhenkanal, Deogarh,Jagatsinghpur, Jajpur, Jharsuguda, Khurda, Koraput, Kalahandi, Kendrapara, Nuapara, Nawarangpur, Nayagarh, Puri, Rayagada, Sundergarh, Sambalpur, Sonepur **- 24 districts**	102	18149	7.99 lakh Ha. of cropped area damaged. Rs.883.42 crores of PU damaged.
18	2003	July-Oct.	Baitarani, Mahanadi, Rushkulya, Vamsadhara, Burhabalanga, Indrabati	Angul, Balasore, Bhadrak, Boudh, Bolangir, Baragarh, Cuttack, Deogarh, Ganjam, Gajapati, Jagatsinghpur, Jajpur, Jharsuguda, Khurda, Koraput, Kalahandi, Keonjhar, Kendrapara, Malkangiri, Nuapara, Nawarangpur, Nayagarh, Puri, Rayagada,Sambalpur, Sonepur **- 26districts.**	92	2956	5.03 lakh Ha. of cropped area damaged. More than Rs.1000.00 crores of PU damaged.
19	2006	July – Aug.	Mahanadi, Brahmani, Baitarani, Subarnarekha, Burhabalanga, Vamsadhara, Rushikulya	Angul, Balasore, Bargarh, Bhadrak, Bolangir, Boudh, Cuttack, Dhenkanal, Gajapati, Ganjam, Jagatsinghpur, Jajpur, Kalahandi, Kandhamal, Kendrapara, Keonjhar, Khurda, Koraput, Malkangiri, Mayurbhanj, Nawarangpur, Nayagarh, Nuapara, Puri, Raygada, Sambalpur, Sonepur **-27 districts.**	90	1656	3.104 Lakh Ha. crop damaged, 0.27 lakh Ha. Sand cast, house damaged 120446 nos, Loss of P.U -Rs.2043.00 crores
20	2007	Jul-Aug.-Sept.		27, 12, 15 districts respectively			
21	2008	Jun–Sept.	Subarnarekha, Burhabalang, Baitarani, Mahanadi, Rushikulya, Vansadhara	Angul, Balasore, Bhadrak, Boudh, Bolangir, Bargarh, Cuttack, Gajpati, Jagatsinghpur, Jajpur, Kendrapara, Khurda, Kalahandi,Keonjhar, Mayurbhanj, Nuapara, Nayagarh, Puri, Rayagada, Sambalpur, Sonepur - **21districts**	110	50163	258155 houses damaged,4.45 lakh Ha. cropped area, 0.14 th Ha sand cast, 651 breaches in rivers, 1276 breaches in canals.

ANNEXURE-II

Standardization of the data

The variables for consideration in clustering are derived by transformation of site characteristics that are measured at different scales. Appropriate transformation by scaling is necessary to ensure that these factors fall between zero and unity (Lim and Lye,2003). Before applying the data in any of the clustering methods the catchments characteristics (variables/attributes of each site) are to be rescaled by the following formula in order to make all values dimensionless (Jingyi and Hall, 2004).

$$x_{kn} = \frac{Y_{kn} - Y_{n(\max)}}{Y_{n(\max)} - Y_{n(\min)}} \quad \text{(A2.1)}$$

Where, Y_{kn} is the n^{th} feature at site k, $Y_{n(\max)}$ and $Y_{n(\min)}$ are the maximum and minimum of the n^{th} feature within the data set. The process is also known as normalization of data.

Hierarchical clustering (HC)

Hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure called a dendrogram. The root of the tree consists of a single cluster containing all observations, and the leaves correspond to individual observations. The dendrogram obtained as an identification of individual sites are tried for different number of clusters. A straight line has been drawn against the similarity measures to show at what similarity measure maximum how many clusters can be formed with how many number of sites.

K-mean method (KM)

This method was developed by MacQueen (1967). It is best described as a partitioning method. It partitions the data into K mutually exclusive clusters and returns a vector of indices indicating to which of the K-clusters it has assigned each observation. The algorithm to clusters N objects based on attributes into K partitions where $K < N$. The objective function is

$$V = \sum_{i=1}^{k} \sum_{x_j \varepsilon S_i} (x_i - \mu)^2 \quad \text{(A2.2)}$$

It tries to achieve minimum intra cluster variance or the squared error function. Where there are K clusters $S_i = 1,2.......K$ and V_i is the centroid or mean point of all the points $x_j \varepsilon S_i$.

Fuzzy C-mean (FCM)

Fuzzy c-means (FCM) is a data clustering technique where in each data point belongs to a cluster to some degree that is specified by a membership grade. This technique was originally introduced by Bezdek (1981) as an improvement on earlier clustering methods. It provides a method that shows how to group data points that populate some multidimensional space into a specific number of different clusters (Fuzzy logic toolbox, MATLAB).

The objective function

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, 1 \le m \le \infty \qquad \text{(A2.3)}$$

Where, m = any real number, u_{ij} degree of membership of x_i in cluster j, x_i = i^{th} of d-dimensional measured data, C_j = d-dimension center of cluster.

Kohonen self organizing feature map (SOM)

Kohonen neural network also known as the self-organizing feature map is a realistic, although very simplified, model of the human brain (Kohonen, 1997). The purpose of the SOM is to capture the topology and probability distribution of input data. Hall and Minns (1999) indicated the feasibility of employing a Kohonen neural network for the classification of hydrological homogeneous regions.

The learning procedure in a Kohonen Map is unsupervised competitive learning. Only the winning node and its neighbors are updated during the learning. Weights w_{ij} are updated using following formula:

$$w_{ij}(new) = w_{ij}(old) + \alpha\left[x_i - w_{ij}(old)\right] \qquad \text{(A2.4)}$$

Where, x_i is the i^{th} input signal, w_{ij} is the weight of the connection from node i to node j and α is the learning rate. The winning node is determined by a similarity measure, which can be Euclidean distance measure or the dot product of two vectors. The Euclidean distance (D_j) that is mostly used for similarity measure is calculated as:

$$D_j = \sqrt{\sum_{i=1}^{n}(x_i - w_{ij})^2} \qquad \text{(A2.5)}$$

Where the Kohonen map based data- clustering technique is applied to show how multi-dimensional datasets can be reduced to 2-D (feature) maps, manifesting clusters of similar data items (Kiang et al., 1997).

Principal component analysis (PCA)

PCA was invented in 1901 by Karl Pearson. It involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

Andrews plot

Andrews (1972) suggested a simple graphical technique to represent a multi-dimensional data by a two-dimensional curve. It provides a good method of viewing patterns of similarity or dissimilarity across multiple dimensions. A point in multi-dimensional space is represented by a curve described by the function:

$$f(t) = \frac{x_1}{\sqrt{2}} + x_2 Sin(t) + x_3 Cos(t) + x_4 Sin(2t) + x_5 Cos(2t) .... \quad (A2.6)$$

Where, x_1, x_2............... are the variables used to characterize a particular site. The function is plotted over the range -π to +π. Curves representing points which are located near one another in multi-dimensional space will look similar, whereas points which are distant will produce different looking curves. Result will depend on the order in which the variables are labeled. The first few variables tend to dominate. So it is a good idea to put the most important variables first.