STATISTICAL MODELING OF RAINFALL CHARACTER OF NORTH-EAST INDIA

A Thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Surobhi Deka
Registration No. 015 of 2010

Department of Mathematical Sciences
School of Science and Technology
Tezpur University, Assam
Tezpur- 784028, India

June, 2010

Abstract

The main objective of this research work is to make a statistical analysis of an important meteorological parameter rainfall with special reference to North East India. More precisely, an attempt has been made to find the best fitting model for the analysis of daily rainfall, sequence of rainfall (i.e. spell) and annual maximum rainfall of North East India.

It is well known fact that Markov chain model can be fitted to daily rainfall occurrence and several authors have used Markov chain model to estimate the wet and dry days in past. First, we demonstrate the application of first order two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons of North East India. Then an analysis regarding the fitting Markov chain of appropriate order has been carried out in this study using the Akaike information criterion. For the majority of the stations Markov chain of order one is identified as the most appropriate model, followed by order two, for describing the daily precipitations occurrences over North East India during Indian summer monsoon season. Then some well known distributions namely, Normal, Log-normal, Gamma and Weibull distribution are also fitted to find the best fitting distribution function to the daily rainfall series. Chi-square test and Kolmogorov-Smirnov test have been performed judging the goodness of fit. Cumulative distribution functions for each of the aforesaid distributions and the observed cumulative distribution functions are plotted for identifying the right probability density function for the daily rainfall amount. The Gamma and Weibull distributions are observed to be competing each other and both are very close to the observed distributions as evinced by the graphical plots.

Again the distribution of rainfall depends on the wet and dry spells over a period of time, so it is desirable to investigate the pattern of occurrence of wet and dry spells especially in Indian summer season (April-September). Various distributions have been fitted to develop a discrete precipitation model for the daily series of precipitation

occurrences over North East India. The goodness of fit of the proposed model have been tested using Kolmogorov-Smirnov test. It is observed that Eggenberger-Polya distribution fairly fits wet and dry spell frequencies and can be used in the future for an estimation of the wet and dry spells in the area under study.

Knowledge of spatial and temporal variability of extreme rainfall events is very much useful for the design of dam and hydrological planning. Therefore, study on the statistical modeling of extreme rainfall is very much essential as the statistical model may vary according to the geographical locations of the area considered. Considerable efforts have been made in this direction using the annual series of maximum daily rainfall data for the period of 42 years of nine stations in North East India. For this purpose, five three-parameter extreme value distributions viz. Generalized Extreme Value distribution (GEV), Generalized Logistic distribution (GLD), Generalized Pareto distribution (GPD), Lognormal distribution (LN3) and Pearson (P3) distribution are considered. The estimation of the parameters for each distribution has been done using the methods of L-Moment and LQ-Moment independently. The performances of the distributions are evaluated using three goodness of fit tests namely relative root mean square error, relative mean absolute error and probability plot correlation coefficient. Further, L-moment ratio diagram is also used to confirm the goodness of fit for the above five distributions. This study reveals that the results of the best fitting distributions may differ for a particular station depending on either L-Moment or LQ-Moment is used. However, generalized logistic distribution is empirically proved to be the most appropriate distribution for describing the annual maximum rainfall series for the majority of the stations in North East India.

Recently, Wang ([73]) introduced the concept of LH-moments as generalization of the L-moment with the capacity of a more detailed analysis of annual flood peak data. These are based on linear combination of higher order statistics. Although a good number of articles is devoted to the statistical modeling of extreme rainfall using L-moments, there is hardly any literature concerning the use of LH-moments in the statistical modeling of extreme rainfall. Therefore, LH-moments(L to L_4) are used to

estimate the parameters of three extreme value distributions viz. Generalized Extreme Value distribution, Generalized Logistic distribution and Generalized Pareto distribution to annual maximum daily rainfall data for the period 1966 to 2007 of nine distantly located stations in North East India. The performances of the distributions are assessed by evaluating the relative bias (RBI'AS) and relative root mean square error (RRMSE) of quantile estimates through Monte Carlo simulations. Then the boxplot is used to show the location of the median and the associated dispersion of the data. This study reveals that generalized Pareto distribution would be appropriate for describing the annual maximum rainfall series in North East India when the distributions are fitted using LH-moments. More precisely, zero level of LH-moments of GPD is found to be more superior to the majority of the stations in comparison to the other higher levels of LH-moments. Further, higher levels of the LH-moments can also be used to obtain improve estimate values of extreme rainfall for some stations in North East India.

Declaration

I, **Surobhi Deka**, hereby declare that the subject matter in this thesis entitled **Statistical Modeling of Rainfall Character of North-East India** is the record of work done by me, that the contents of this thesis did not form basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the thesis has not been submitted by me for any research degree in any other university/institute.

This thesis is being submitted to the Tezpur University for the degree of Doctor of Philosophy in Mathematical Sciences.

Place: Napaam

Date: 29/06/2010

Surobhi Deka.

(Surobhi Deka)

TEZPUR UNIVERSITY

Certificate

This is to certify that the thesis entitled **Statistical Modeling of Rainfall Character of North-East India** submitted to the School of **Science and Technology** Tezpur University in partial fulfillment for the award of the degree of Doctor of Philosophy in **Mathematical Sciences** is a record of research work carried out by Ms **Surobhi Deka** under our supervision and guidance.

All help received by her from various sources have been duly acknowledged. No part of this thesis has been submitted elsewhere for award of any other degree.

Prof. Munindra Borah
Department of Mathematical Sciences
Tezpur University, Tezpur, Assam
Tezpur-784028

Prof. Sarat Chandra Kakaty
Department of Statistics
Dibrugarh University, Dibrugarh, Assam
Dibrugarh-786004

Acknowledgement

I am glad to take this opportunity to express my deep sense of gratitude to Prof. Munindra Borah who introduced me to this research field and whose guidance has enabled me to present this thesis. I am also thankful to him for making me feel free to express my views and for sharing an excellent rapport with me.

Also I express my special appreciation to Prof. Sarat Chandra Kakaty, Department of Statistics, Dibrugarh University, for agreeing to be my co-supervisor. I am deeply indebted to him for his invaluable guidance, immense patience, utmost care and constant encouragement through out this work.

I owe my thankfulness to Prof. N. D. Baruah, Head, Department of Mathematical Sciences, Tezpur University, for providing me with the necessary facilities required for my work. I also take this opportunity to express my gratitude to the esteemed faculty members of the Department of Mathematical Sciences for their help on various occasions. I also extend my thank to our system analysts Anajan da, Naba, Debajit da and Biswajit da for their help in the computer related works.

I thank Dr. Jiten Hazarika, Reader, Department of Statistics, Dibrugarh University for his cooperation and useful comments during the progress of my PhD work.

This work would have not been possible but for the strong moral support and the unfailing faith of my husband that has sustained me throughout this endeavour. I have no words to express my thankfulness to him. I owe my appreciation to all my family members for their encouragement during the course of my work, especially to my sister Jeemoni for her kind presence and moral support during my pathetic health conditions.

I thank the Regional Meteorological Centre, Guwahati and the Tocklai Experimental Station, Jorhat for providing the data needed for this work. In this context, I would also like to thank my friends Barnali, Dipa, Mousumi and Jolly for their kind help and cooperation during the period of data collection.

I thank the Council for Scientific and Industrial Research, India for providing financial support at the start of this work and the Department of Science and Technology, India towards the end of this PhD work.

Last but not least, I must say that without the love and encouragement of all of my friends at Tezpur University, this investigation would not have been possible. All of my friends at Tezpur University contributed in their own way. I must thank Nandini ba, Kanan, Pubali, Mandakini, Jonali, Chumchum, Padmakshi, Ajanta ba, Abhijeet, Tazzudin, Anupam, Narayan and Bimalendu with whom the time spent in the last five years was enjoyable and memorable. Thanks are also due to Indranil da and Jagat da for their friendship and help in all official matters. Thus, through this acknowledgement I express my heartiest regards to all of my well-wishers.

June, 2010

Truly Yours

(Surobhi Deka)

Department of Mathematical Sciences

Tezpur University

Contents

List of Tables

List of Figures

Chapter 1

General Introduction

The severity of the weather, which manifests in the form of floods and landslides on account of rainfall, has a substantial impact on the life and properties. Rainfall is one of the fundamental components of the hydrological cycle as its accurate estimation is necessary for planning, designing and operation of water resources development programmes. The purpose of this thesis is to find the best fitting statistical model(s) for the analysis of rainfall data of North East India. Fitting a probability distribution function to observed data provides a compact and smoothed representation of the frequency distribution revealed by the available data, and leads to a systematic procedure for extrapolation to frequencies beyond the range of the data set (Stedinger et al. [68]).

1.1 Background

We live in a world that is exposed to the vagaries of severe and unusual weather. Natural disaster and severe weather events have a close link because all severe weather events, due to climate change or otherwise could and often lead to natural disasters that occur on varying time and space scales. Disaster may strike any country but the greatest burden falls on less developed countries and their highly populated regions. Despite development in all fields of socio-economic activity we have not succeeded in insulating the population from their effects. One of the common features of developing countries in the South East Asia is flash flood in urban areas during rainy season and acute shortage of water for domestic and agriculture uses during winter.

North East India, located at east of 80^{o} E and North of 21^{o} N, is one of the major disaster prone region of India because of their unique geographical locations and physical features, witnessing the fury of monsoon. The summer monsoon influence this region from June to September contributing more than 80% of the annual rainfall. During this season major floods occur that often lead to disaster. The average annual rainfall in North East India ranges from 2000-4000 mm with a maximum of 11000 mm in Cherrapunjee. However, more than the total amount, the distribution of rainfall matters a lot for sustained high yield of agricultural crops throughout the season. In the North East India, the rainfall distribution is not even. While the excess rainfall in the monsoon months of June-September causes drainage problems, in the longer dry spell during November to March crop goes down in spite of having sufficient rainfall in the monsoons. Again it is important to note that the distribution of water resources potential in the country shows that as against the national per capita annual availability of water as 2208 m^3 the average availability in North East because of the Brahmaputra and the Barak rivers is as high as 16589 m^3. However, this vast water resource remains unutilized and creates problems in the entire region in many ways. This necessitates changes in perspective of water management in the region.

Extreme precipitation events (heavy rain storm, cloud burst) may have their own impacts on the fragile geomorphology of the Himalayan part of the Brahmaputra basin causing more widespread landslides and soil erosion. The response of hydrologic systems, erosion processes, and sedimentation in the Himalayan river basins could alter significantly due to climate change. Two extremely intense cloud bursts of unprecedented intensity- one in the western Meghalaya hills and Western Arunachal Pradesh in 2004 produced two devastating flash floods in the Goalpara and Sonitpur districts of Assam bordering Meghalaya and Arunachal respectively causing hundreds of deaths. The most recent examples of such flash floods originating from extreme rainfall are two events that occurred in the north bank of the Brahmaputra River and caused significant damage to human life and property. The first of the two events occurred during the monsoon season on June 14th, 2008 due to heavy rainfall on the hills of Arunachal Pradesh north of Lakhimpur District causing flash floods in the rivers of Ranganadi, Singara, Dikrong and Kakoi that killed at least 20 people and inundated more than 50 villages leading to displacement of more than 10,000 people. The other that occurred in the post monsoon season on October 26 affected a long strip of area of northern Assam valley adjoining foothills of Bhutan and Arunachal Pradesh causing flash flooding in four major rivers (all are tributaries of the river Brahmaputra) and a number of smaller rivers. This episode of flash floods caused by heavy downpour originated from the Tropical Depression 'Rashmi', (a depression over the West Central Bay of Bengal adjoining Andhra coast).

Climate change has cascading and far reaching affects on almost every aspect of environment and societies as already observed amply all over the world. The developing countries of the world with large populations living in poverty and degraded environments and reliant on primary production are most vulnerable to the impacts of global climate change. The northeast Indian region of India is expected to be highly prone to the consequences to climate change because of its geo-ecological fragility, strategic location vis-à-vis the eastern Himalayan landscape and international borders, its trans-boundary river basins and its inherent socio-economic instabilities. Environmental security and sustainability of the region are and will be greatly challenged by these impacts.

Studies on rainfall and the temperature regimes of northeast India indicate that there is no significant trend in rainfall for the region as a whole i.e. rainfall is neither increasing nor decreasing appreciably for the region (cf. [13, 29]). However, for a part of the region that the meteorologists of the country officially refer to as the 'South Assam Meteorological Subdivision'(that covers mainly the hill states of Nagaland, Manipur, Mizoram and Tripura and parts of the Barail Hills in southern Assam), a significant change in seasonal rainfall has been observed. The summer monsoon rainfall is found to be decreasing over this region significantly during the last century at an approximate rate of 11 mm per decade (cf. [13, 48]). For example, several districts of Assam were badly affected due to drought like situations consecutively for two years in 2005 and 2006 which had a signature of climate change on them as vindicated by the Intergovernmental Panel on Climate Change(IPCC) report of 2007([35]). In the intense drought-like conditions that prevailed in as many as 15 districts of Assam during the summer monsoon months of the year 2006 owing mainly to below normal (nearly 40%) rainfall in the region, more than 75% of the 26 million people associated with livelihoods related to agriculture in these districts were affected and the state suffered a loss of more than 100 crores due to crop failure and other peripheral affects. Other states of the region also received rainfall 30-40% below their normal rainfall except Mizoram. Normally such fluctuations are considered as results of inter-annual variability of the monsoons, but then climate change impacts are supposed to affect the southwest monsoon also by increasing the normal mode of its variability.

Rainfall occurring earlier or later has adversely affected sowing and harvesting of crops. Moreover, there are reports that natural wetlands are shrinking in many parts of the region. Some ecologists have informed about appearance of more number of invasive species and changes in their distribution pattern in the region. Some have reported more number of diseases and pests in citrus species. One significant impact which many plant scientists agree to is the change taking place in the phenological phases in plants ([34]). Besides such scientific evidences, which are however few, individual and collective opinions in various parts of the region bear references to what may be construed as increased variability or changes in local climates. Such anecdotal references talk of irregular rainfall pattern with rainfall starting quite early in the region (say in January), heavy rainfall events (extreme rainfall) and flash floods becoming more frequent and dry periods becoming longer in various parts of the region. However, more rigorous study on the rainfall character of North East region of India needs to be done at regional scale before anything can be said conclusively.

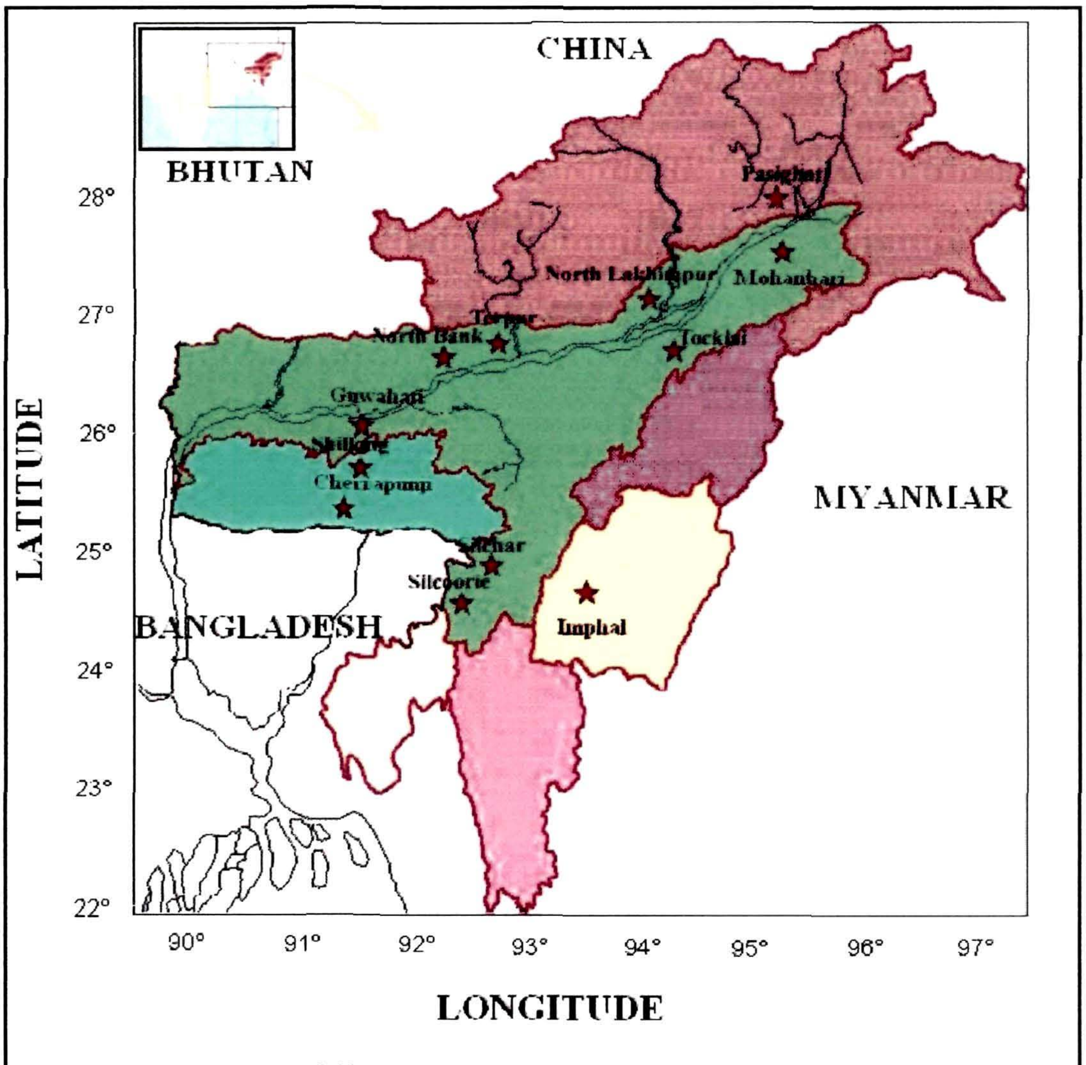

Figure 1.1 Locations of rain gauge staions used in this study

1.2 Brief Description of the Study Area

The brief description, especially hydroclimatology, of rain gauge stations used in this study is presented in this section. The geographical locations of the rain gauge stations are shown in Figure 1.1.

Guwahati: Guwahati is the largest city in the North East Region of India and is located at 26°11′N 91°44′E. Guwahati's climate is mildly sub-tropical with warm, dry summers from April to late May, a strong monsoon from June to September and cool, dry winters from late October to March.The city experiences an annual rainfall of 180 cm (from May to September) with an average number of 77.3 rainy days. While summer temperatures range from 22°C to 38°C, in winters the mercury ranges from 10°C to 25°C.

Shillong: Shillong is the capital of Meghalaya, one of the smallest states in India. Shillong is located at 25°34′N 91°53′E. It is on the Shillong Plateau, the only major pop-up structure in the northern Indian shield. Due to its latitude and high elevation Shillong has a sub-tropical climate with mild summers and chilly to cold winters. Shillong is a subject to vagaries of the monsoon. The monsoons arrive in June and it rains almost until the end of August. In summers the average temperature is 23 degree Celsius and in winters it is dropped to 4 degree Celsius.

Cherrapunji: Cherrapunji is the world's wettest place and is just 56 km from the capital Shillong of Meghalaya. Geographically it is located at 25°18′N 91°42′E. Cherrapunji's yearly rainfall average stands at 11,430 mm. This figure places it behind only nearby Mawsynram, Meghalaya, whose average is 11,873 mm and Mount Waialeale (USA) on the Hawaiian island of Kauai, whose average is 11,684 mm. The orography of the hills with many deep valleys channels the low flying (150-300 m) moisture laden clouds from a wide area to converge over Cherrapunjee which falls in the middle of the path of this stream. The winds push the rain clouds through these gorges and up the steep slopes. The rapid ascendance of the clouds into the upper atmosphere hastens the cooling and helps vapours to condense. Most of Cherrap of air being lifted as a large body of water vapour. Extremely large amount of rainfall at Cherrapunjee is perhaps the most well known feature of orographic rain in northeast India.

Imphal: Imphal is the capital of Manipur, located at 24°49′N 93°57′E. It has an average elevation of 786 metres (2578 feet). It is located in the extreme east of India. The Imphal Valley is drained by several small rivers originating from the hills surrounding it. Imphal has a sub-tropical climate with cool, dry winters, a warm summer and a moderate monsoon season. July is the hottest month with temperatures averaging around 25°C, while January is the coldest with average lows near 4°C. The city gets about 1320 mm of rain with June being the wettest month.

Mohanbari: Mohanbari is located 15 km from the city center Dibrugarh of district Dibrugarh, Assam, India. Being located 27°26′60N 95°1′60E and with its unique physiographic elements, the area experiences subtropical monsoon climate with mild winter, warm and humid summer. Rainfall decreases from south to north and east to west in the area. The average annual rainfall of the Dibrugarh city in the north is 276 cm with a total number of 193 rainy days, while at Naharkatia in the south, it is 163 cm with 147 rainy days.

North Lakhimpur: North Lakhimpur is situated in the eastern parts of India in the state of Assam. The district of Lakhimpur lies on north bank of the mighty river Brahmaputra. It is situated at 27°13′60 N and 94°7′E.

Pasighat: Pasighat is the headquarter of East Siang district in the Indian state of Arunachal Pradesh and located at 28.07°N 95.33°E. It has an average elevation of 153 metres. The area experience tropical humid climate during summer and dry mild winter. The place is known for receiving highest rainfall in a single year. In fact Pasighat and area around it receive heavy rainfall every year during monsoon season starting from May till September.

Silchar: Silchar is the headquarter of Cachar district in the southern part of the Assam state and it lies between latitude 90.44 E and longitude 20.04 N. Because of its subtropical monsoon climate, silchar experiences high rainfall, about 85 % of which occurs during May to October. The average annual rainfall in Silchar varies from 2500 mm to 3400 mm and the temperature is moderate ranging from 13°C-35°C.

Tezpur: Tezpur is situated in the eastern parts of India in the state of Assam and located at 26.63°N 92.8°E. It has an average elevation of 157 ft and the average annual rainfall in Tezpur ranges from 2000 mm to 2700 mm. The climate in this part of Assam is usually pleasant, the only problem arises due to the high humidity factor. The summer see the temperature rising as high 34.6°C and during winters the temperatures may drop to about 12°C.

Tocklai: Tocklai is situated in the district Jorhat, Assam, India and its geographical coordinates are 26°45′ N 94°13′ E and 91 meter above mean sea level. The average annual rainfall in this part of central Assam ranges from 2000-3000 mm. This place is also known for Tocklai Experimental Station which has been serving the tea industry and has become synonymous with the research on tea in the country.

Silcoorie: Silcoorie is just 10 km from the district headquarters Silchar of Cachar district and is located at 24°50′ N 92°48′ E. The average annual rainfall in Silcoorie varies from 1965 mm to 3000 mm.

North Bank: North Bank station is 35 kms from Tezpur town on the north bank of the River Brahmaputra in Assam and located at 26°50′ N 92°38′ E.

1.3 Motivation and Objectives

This section elucidates our main objectives and motivation for the present study. Realistic sequences of meteorological variables such as precipitation are key inputs in many hydrologic, ecologic and agricultural models. Rainfall information form the basis for designing water related structure in agriculture planning, in weather modification, in water management and also in monitoring climate changes. The most commonly measured and recorded information on rainfall is a daily value gauged. The equipments for observing daily values are also the simplest type of rain gauges which are fairly inexpensive, easy to maintain and read by local observer with little expertise.

In India majority of the people depends on agriculture for their livelihood. And Indian agriculture primarily depends on rainfall. Only 20% of the cultivated land enjoys the facilities of irrigation. In remaining areas, however the farming is done under unirrigated conditions and as such it, depends mostly on the occurrence of rainfall. Agriculture is highly sensitive to rainfall modulation during the rainy season/Indian Summer Monsoon season which provides more than 80% of the annual rainfall over India. Therefore, a detailed knowledge of rainfall regime is an important prerequisite for agriculture planning.

Amount of daily rainfall is an important factor that impacts agriculture system. It governs the crop yields and determines the choice of the crops that can be grown. Therefore it is also important to graduate the rainfall of different time scales by fitting appropriate frequency distributions. According to Fisher [21] crop yield during a season is mainly influenced by the distribution of rainfall rather than season total amount of rainfall. Again the distribution of rainfall depends on the wet and dry spells over a period of time. So it is of the essence to investigate the pattern of occurrence of wet and dry spells during especially in Indian summer monsoon season (June-Sept.). The occurrence of wet and dry spells can be regarded as series of Bernoulli trials. And so, the pattern of occurrrence of rainfall can be investigated by fitting a stochastic model to rainfall data over a moderate period of time(summer monsoon) during which agriculture operation is highly influenced by rainfall.

Another important issue which we need to address is the extreme rainfall. Extreme rainfall events can have severe impacts on society. It afflicts the worst environmentally related tragedy, which contributes to loss of crops and valuable property and untold human misery. The vulnerability of the people to the extreme weather events seems to be increasing every year in terms of change in frequency and adversely affecting the people. Stochastic models for extreme rainfall events over an area may be used for such disaster prevention purposes. If the best fitting distribution is known for a particular station, one would be able to predict the return value of this extreme rainfall event at a specific time in the future. The main objectives of this thesis are outlined as follows:

- **Statistical modeling of the pattern of occurrence of daily rainfall data**

  Here we have made an effort to demonstrate the application of first order two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons in North East India. Further an analysis regarding the fitting of Markov chain of appropriate order has been carried out in this study.

- **Application of well known probability distributions**

  Some well known probability distributions viz. normal, log-normal, gamma and Weibull distribution have been fitted to find the best fitting distribution function to the daily rainfall series in North East Region of India. Chi-square test and Kolmogorov-Smirnov test have been performed to judge the goodness of fit.

- **Statistical modeling of wet and dry spell frequencies**

  In this study, the point of approach has been taken as the modeling of the duration of consecutive dry and wet days i.e. spell, instead of individual wet and dry days. Various distributions have been fitted to describe the wet and dry spell frequencies of occurrences considering the climatic features of the different parts of North-East India.

- **Statistical analysis of annual maximum rainfall**

  Knowledge of extreme rainfall event is very much useful for the design of dam and hydrological planning. Again the statistical model may vary according to the geographical locations of the area considered. Therefore, it is very much essential to make a study on extreme rainfall over North East India. For this purpose, several extreme value distributions have been fitted. The estimation of the parameters for each distribution has been done by using the methods of L-moment, LQ-moment and LH-moments independently.

1.4 A Brief Survey on Statistical Methods

Nature is a very complicated system. Despite this, it is often necessary to simulate the behavior of natural systems. To attempt this by modeling all the physical processes deterministically is a very difficult task. Instead, stochastic models are advocated. Stochastic models generally have a relatively simple framework that incorporates an element of uncertainty in the outcome. This randomness or uncertainty represents the part of the process that can not be explained deterministically. Stochastic models are designed to reproduce the important patterns evident in the observations based on the current knowledge of the physical processes.

Stochastic modeling of rainfall data has become a frontier research area over the years. The majority of stochastic models deal with either daily rainfall or series of rainfall (i.e. spell) or annual maximum rainfall. In order to put our discussion into proper perspective, we first give a brief account of the development of the statistical modeling of rainfall data over the globe.

As far our knowledge is concerned, the statistical modeling of rainfall data started with the work of Gabriel and Neuman [22, 23]. They applied a first order chain to Tel Aviv precipitation data on the basis of multiple hypotheses testing procedures and it is observed from their study that two state Markov chain give a good description of the occurrences of wet and dry days during the rainy period at Tel Aviv. Bhargava *et al.* [9] studied the occurrence of rainfall with the help of Markov chain model of order one in Raipur District India. Further, the occurrences of wet and dry weeks were studied by Gore and Thapliyal [28] at Maharastra, India. Latter, Gates and Tong [28] reanalyzed the same Tel Aviv data applying the AIC procedure and suggested that a Markov chain of order not lower than 2 should be fitted, instead of the previously fitted first order. Although there is a disagreement on the appropriate order for the Tel Aviv model but one must agree that Markov chains are obvious candidate to model the occurrence of rainfall. Some authors attempted to describe rainfall amounts by fitting Markov chains with many states each representing a range of amounts. One unsatisfactory element of these models has been the large number of parameters to be estimated. Then some well known distributions were tested as an alternative to Markov chains with many states to estimate the amounts of rainfall. Barger and Thom [6] showed that gamma distribution provides good fit to precipitation series in the United States. The best-fit gamma distribution was also found by Simpson [67] based on rather evaluated rainfall data. Mooley [49] tested whether a suitable unified probability model exists or not for the distribution of monthly rainfall associated with the Asian Summer Monsoon. He found that gamma distribution is the most suitable probability model from among the Pearsonian models. Gamma distribution was also fitted by Stern and Coe [69] for modeling rainfall amount. It was claimed that a comprehensive analysis of rainfall data should use daily records and not based on 7, 10 days or monthly total. Sharma [65] claimed that the probability estimation for the Weibull pdf can be done by analytical integration which was not possible for normal, lognormal and gamma probability distributions. Aksoy [5] investigated the amounts of daily rainfall and the ascension curve of the hydrograph by using 2-parameter gamma distribution. Muralidharan and Lathika [51] analyzed the rainfall occurrence based on modified version of Weibull distribution for two meteorological stations in India.

Let us turn our discussion to the statistical modeling of wet and dry spells over different parts of the world. The most frequently used model for generating consecutive sequence of dry and wet days is the first order, two state, homogeneous Markov chain that has been applied by several authors (cf. Gabriel et al. [22], Katz [36], Bruhn et al. [10], Richardson [61], Geng [25], Matyasovszky et al. [43], Wilks [75], Dubrovsky [20]). The major disadvantage of this model is that it overestimates the very short, but underestimates the very long dry sequences. An essential improvement to reproduce the short and long spells were made by Berger et al. [8] and Nobilis [54] using higher order Markov chain and Eggenberger-Polya distribution. They found that short spells were best fitted by fourth order Markov chain, where as the Eggenberger-Polya distribution gave the best fit to the long series. Later, Racsko et al. [58] proposed a model constituting two different geometric distributions. In the referred study, both the geometric distributions were separated according to the length of dry spells. Results of the works suggested that mixed distribution, including geometric one, could be promising in reproduction of long dry periods (where as simple geometric distribution gave the best fit for wet spells). For wet spells, it was also observed that simple geometric distribution could be promising. Recently, following the idea of [58] a mixture distribution based on a weighted sum of two geometric distributions, as well as that of one geometric and one poisson distribution was applied by Wantuch et al. [74]. The first model exhibits good fitting for the dry spells and the latter one can be advised to employ for the wet periods. More recently, while Tolika et al. [70] found that both Markov chain of order two and negative binomial distribution can be used to estimate the wet spells in Greece, Eggenberger-Polya and truncated negative binomial were found to be more efficient in fitting observed data both for wet/dry spells by Giuseppe et al. [26].

Applications of extreme value distributions to rainfall data have been investigated by several authors from different regions of the world. Rakhecha et al. [59] analyzed the annual extreme rainfall series at 316 stations over the Indian region, covering 80 years of rainfall data for trend and persistence using standard statistical tests. For investigating more generalized issues regarding the adequacy of extreme value distributions for extreme rainfall analysis, Baloutsos et al. [7] made the statistical analysis for the longest rainfall record available in Greece. In the same direction, Koutsoyiannis [41] made an extensive empirical investigation of the longest available rainfall records worldwide, each having 100-154 years of data. Nadarajah et al. [52] and Nadarajah [53] provided the application of extreme value distributions to rainfall data over sixteen locations spread throughout New Zealand and fourteen locations in West Central Florida, respectively. Extreme value distributions were also used by Aronica et al. [1] to analyze the trend in the extreme rainfall series for a fixed return period by estimating the maximum rainfall depth in Palermo, Sicily, Italy. They estimated the parameters using L-moments. Zalina et al. [76] discussed the comparative assessment of eight candidate distributions in providing accurate and reliable maximum rainfall estimates for Malaysia. Model parameters were estimated using the L-moment method. They concluded that the GEV distribution is the most appropriate distribution for describing the annual maximum rainfall series in Malaysia. On the other hand, Zin et al. [78] found GLD as the most frequently selected best fitting distribution and LN3 as the least frequently selected distribution for extreme rainfall in Peninsular Malaysia. Those results differ from the results obtained by Zalina et al. in [76]. While Kotz et al. [40] made an extensive study on the fitting of extreme value distribution, the detailed references on the statistical modeling of annual maximum rainfall based on L-moment and LQ-moment can be found in [78]. The most commonly used distributions for extreme rainfall data can be found from the references such as Hosking and Wallis [32] and Rao and Hamed [60]. Recently, Wang ([73]) developed the LH-moments as a generalization of the L-moments with the capacity of a more detailed analysis of annual flood peak data. In his study he concentrated only on the generalized extreme value distribution. Since then LH-moments have been used by several authors in flood frequency analysis. Meshgi et al. ([46], [47]) performed a comparative study of L and LH-moments for regional flood frequency analysis of Kharkhe watershed, located in Western Iran. In their study, they extended the regional homogeneity test for L-moment developed by Hosking ([31]) to each LH-moments level from L_1 to L_4 and also developed the LH-moments for generalized logistic distribution (GLD) and generalized Pareto distribution (GPD). The other application of LH-moments is due to Hewa et al. ([33]) in low flow frequency analysis. They developed a method based on LH-moment to use GEV distribution to model the lower tail of low-flow frequency curve, without explicitly censoring the data sample. It is observed from their analysis that GEV/LH-moment method is more suitable method to model low flows.

1.5 Organization of the Thesis

The thesis consists of seven chapters followed by appendices and bibliography. The organization of the thesis is as follows: Chapter 2 deals with the the application of first order two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons in Imphal, Mohanbari, Guwahati and Cherrapunji. The study reveals that the occurrence of wet and dry days in the tract can be rightly described by a two state Markov chain. Further, it is observed that the number of wet days varies from 47 to 65 for Imphal, Mohanbari and Guwahati and 96 for Cherrapunji.

Chapter 3 demonstrates the application of the Akaike information criterion to determine the order of two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons in North-East India. First order Markov chain model has been found to be an adequate model for most of the stations of North East regions of India to determine the daily precipitation.

In Chapter 4, an attempt has been made to examine the goodness of fit of some well known probability distributions based on daily rainfall observations sampled from seven distantly located stations in North East Region of India Viz.Imphal, Mohanbari, Guwahati, Cherrapunji, Silcoorie, North Bank, Tocklai (Jorhat). The gamma and weibull distributions are observed to be competing each other and both are very close to the observed distributions as evinced by the graphical plots.

Chapter 5 is concerned with modeling of duration of consecutive dry and wet days i.e. spell, instead of individual wet and dry days. Various distributions have been fitted to describe the wet and dry spell frequencies of occurrences. The goodness of fit of the proposed models have been tested using Kolmogorov-Smirnov test. It is observed that Eggenberger-Polya distribution fairly fits wet and dry spell frequencies and can be used in the future for an estimation of the wet and dry spells in the area under study.

Considerable efforts have been made in Chapter 6 to determine the best fitting extreme value distribution to describe the annual series of maximum daily rainfall data for the period 1966 to 2007 of nine distantly located stations in North East India. Model parameters are estimated using the method of L-moment and LQ-moment. This study reveals that the results of the best fitting distributions may differ for a particular station depending on either L-moment or LQ-moment is used. However, generalized logistic distribution is found to be more consistent in comparison to the other three best fitting distributions.

Chapter 7 is concerned with the application of LH-moments as generalization of the L-moment to describe the annual series of maximum daily rainfall data. LH-moments(L to L_4) are used to estimate the parameters of three extreme value distributions viz. generalized extreme value distribution, generalized logistic distribution and generalized Pareto distribution to annual maximum daily rainfall data for the period 1966 to 2007 of nine distantly located stations in North East India. The performances of the distributions are assessed by evaluating the relative bias (RBIAS) and relative root mean square error (RRMSE) of quantile estimates through Monte Carlo simulations. Then the boxplot is used to show the location of the median and the associated dispersion of the data. Generalized Pareto distribution has been found to be appropriate to the majority of the stations for describing the annual maximum rainfall series in North East India using LH-moments.

Some chapter-wise information/results in the form of Tables, which are integral and underpin the findings of the chapters are finally appended at the end.

Chapter 2

Statistical Modeling of Daily Rainfall Data: Markov Chain Approach

In this chapter, we demonstrate the application of first order two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons of North East India. The study reveals that Markov chain model can be used to study the daily rainfall occurrence of North East region of India.

2.1 Introduction

In India, climate is considered to be one of the major constraint of agriculture and agricultural planning, and agriculture scientists/policymakers often keep this in mind while making agricultural plans and policy decisions. Although the various climatic variables interact with the crop in complex ways, rainfall is the limiting factor in most part of the tropics. A fore knowledge of rainfall pattern is of immense help not only to farmers, but also to the authorities concerned with planning of irrigation schemes. The present study is an effort to demonstrate the application of first order two state

Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons in Imphal, Mohanbari, Guwahati and Cherrapunji. The selected stations Imphal, Mohanbari and Guwahati are classified as moderate rainfall area whereas Cherrapunji is classified as heavy rainfall area. The study utilizes five years (2001-2005) daily rainfall data in mm for the summer monsoon months of June, July, August and September. The statistical modeling of daily rainfall data based on first order two state Markov chain can be found in [9], [22], [23] and [28].

The organization of this chapter is as follows. In Section 2.2, we describe the analytical procedure which is needed for the present study. Section 2.3 is concerned with the results and discussion, and finally, the chapter ends with a concluding remark in Section 2.4.

2.2 Analytical Procedure

Any season of the year can be defined as a sequence of wet and dry days, assuming that the occurrence of rain in any day depends only on the occurrence of rain on the previous day, the following conditional probabilities can be defined

$$\begin{aligned} p_1 &= P_r\{\text{wet day/ previous day was wet}\} \\ p_0 &= P_r\{\text{wet day/ previous day was dry}\}. \end{aligned}$$

Thus a hydrological system can be described by two possible states, the dry state and the wet state. The transition probability matrix from one state to another has the following form

state	dry	wet
dry	$1 - p_0$	p_0
wet	$1 - p_1$	p_1

Based on the daily rainfall during the period 1^{st} June to 30^{th} September in each year and for each station, each day is classified as dry day if the amount of rainfall is less than 3 mm and wet day if the amount of rainfall is greater than or equal to 3 mm. Assuming that the occurrence of rainfall on the 1^{st} June depends on the occurrence of rainfall on 31^{st} May and repeating this process for each year the transition count for each possibilities can be calculated. Let these be denoted by n_{00}, n_{01}, n_{10}, n_{11} where $n_{00} + n_{01} = n_0$ and $n_{10} + n_{11} = n_1$.

	dry	wet	total
dry	n_{00}	n_{01}	n_0
wet	n_{10}	n_{11}	n_1

The two parameters p_0 and p_1 are required to be estimated for describing the Markov chain. The maximum likelihood estimates of p_0 and p_1 are given by

$$p_0 = \frac{n_{01}}{n_0} \quad \text{and} \quad p_1 = \frac{n_{11}}{n_1}$$

with variances of the estimates as

$$\frac{\hat{p}_0(1-\hat{p}_0)}{n_0} \quad \text{and} \quad \frac{\hat{p}_1(1-\hat{p}_1)}{n_1}, \text{ respectively.}$$

In order to demonstrate that the occurrence of a wet day (or dry day) is influenced by the immediately preceding day's weather, we compute the usual normal deviate test statistic

$$Z = \frac{p_0 - p_1}{\text{S.E. of } (p_0 - p_1)}.$$

When the occurrence of a wet or dry day is influenced by the previous day's weather, the above process of the occurrence of wet and dry days over a given time is strictly a two state Markov chain with four transition probabilities, depending on only two parameters as described above.

To obtain the common estimates of these two parameters pooled over all such stations let the four cell frequencies for the i th centre be denoted by n_{00i}, n_{01i}, n_{10i}, n_{11i}, respectively with $n_{00i} + n_{01i} = n_{0i}$ and $n_{10i} + n_{11i} = n_{1i}$. The common estimates of p_0 and p_1 pooled over all the stations are given by

$$\bar{p}_0 = \frac{\sum_i n_{01i}}{\sum_i n_{0i}}, \quad \bar{p}_1 = \frac{\sum_i n_{11i}}{\sum_i n_{1i}}.$$

Considering these common estimates as the expected frequencies at each of the stations two chi-squares for each stations can be calculated for testing the discrepancies between the observation and the expectation. The two chi-squares for the ith centre with 1 degree of freedom are given by

$$\chi^2_{p_0} = \frac{n^2_{00i}}{n_{0i}(1-\bar{p}_0)} + \frac{n^2_{01i}}{n_{0i}\bar{p}_0} - n_{0i},$$
$$\chi^2_{p_1} = \frac{n^2_{10i}}{n_{1i}(1-\bar{p}_1)} + \frac{n^2_{11i}}{n_{1i}\bar{p}_1} - n_{1i}.$$

If any stations show insignificant chi square values for both the parameters then it can be regarded as similar in the pattern of the occurrence of rainfall. They can therefore be grouped together for obtaining common estimates of the two parameters in the usual manner.

From the properties of first order two state Markov chain after a sufficiently long period of time, the system settles down to a condition of statistical equilibrium in which the state occupation probabilities are independent of initial conditions. If this is so then there is an equilibrium probability distribution $\pi = (\pi_0, \pi_1)$. Theoretically this probabilities can be calculated by

$$\pi_0 = \frac{1-p_1}{1-p_1+p_0},$$
$$\pi_1 = \frac{1-p_0}{1-p_1+p_0}.$$

Further quantity of interest is the distribution of the number of successes in a sequence of dependent Bernoulli trials as discussed in Cox and Millar [12]. If Y_n is the number of wet days out of n then Y_n behaves like a sum of independent random variables and asymptotically follows normal distribution with

$$E(Y_n) \backsim \frac{np_0}{1-p_1+p_0}$$
$$V(Y_n) \backsim \frac{np_0(1-p_1)(1+p_1-p_0)}{(1-p_1+p_0)^3}.$$

The properties of the distributions of the length of wet and dry spells in two state Markov chain rainfall model are given in Cox and Millar [12]. A wet spells of length

W is defined as W successive wet days followed by a dry day. The probability that W takes a specific value n is given by

$$P(W = n) = (1 - p_1)p_1^{n-1}$$

which follows geometric distribution. The expected length of wet spells is then given by

$$E(W) = \frac{1}{1 - p_1}.$$

Similarly for the length D of dry spell,

$$\begin{aligned} P(W = n) &= p_0(1 - p_0)^{n-1} \\ E(W) &= \frac{1}{p_0}. \end{aligned}$$

A weather cycle may be defined as a wet spell followed by a dry spell. The distribution of the length C of a cycle is therefore the convolution of two independent geometric distributions. Hence

$$E(C) = E(W) + E(D).$$

2.3 Results and Discussions

The transition counts n_{00}, n_{01}, n_{10}, n_{11} were calculated for each station separately in Table 2.1 and then the transition probabilities p_0, p_1 were estimated using the formulae cited above, and the values for p_0, p_1 are presented in Table 2.2. As a first step to fit a Markov chain model to the data, the difference in the estimates of these two probabilities were tested for significance in respect to each station by usual normal deviate test. The value of $|Z|$ for the stations Imphal, Mohanbari, Guwahati and Cherrapunji were found to be 4.479, 7.472, 4.704 and 7.605, respectively. Thus, the differences are found to be highly significant for all the stations. This shows that the weather of a day is influenced by the weather of the previous day. As such the occurrences of wet and dry days in the tract can be rightly described by a two state Markov chain.

To obtain the common estimates of the two parameters pooled over such stations as are homogeneous for them, χ^2 tests were done by using the formulae given in the Section 2.2. It was seen that for all the stations, χ^2 values were significant for both the parameters which indicates that the occurrence of rainfall varies from station to station. So these stations cannot be grouped together to get the common estimates. The different values of tabulated and calculated χ^2 are given in the Table 2.3.

The various properties of the Markov chain as explained in Section 2.2 were obtained and are given in the Table 2.4. From the Table 2.4 it is seen that, for the first three stations i.e. for Imphal, Mohanbari and Guwahati the expected length of the wet spell varies between 2.00 to 3.03 and for the station Cherrapunji its length is 7.14 i.e.,for the first three stations after two to three consecutive wet days a dry day is expected to occurs but in Cherrapunji a dry day is expected to occurs after 7 or 8 days. Similarly the expected length of dry spells varies between 2.63 to 3.23 for the first three stations and for Cherrapunji its value is 1.96. So in Cherrapunji after 1 or 2 consecutive dry days a wet day is expected to occur and for the remaining three stations after three consecutive dry days a wet day is expected. The expected days of the cycle is therefore 5 to 6 days for the first three stations and 9 for cherrapunji.

The expected number of wet days during the period of 122 days and the actual number of wet days during that period are shown in columns 8 and 9 respectively of Table 2.4. The results in both the columns are almost same and for the first three stations its values vary between 47 to 65 days and for Cherrapunji its value is 96 days. The standard deviation of the distribution is about 7 days.

The state occupation probabilities at equilibrium i.e. π_0 and π_1 which are independent of the initial conditions and the number of days required to get the state of equilibrium were also obtained and are given in the column 12, 13, 14 of the Table 2.4. It is seen that the number of days to equilibrium varies from 6 to 9 which shows that after 6 to 9 days from 1^{st} June, the state occupation probabilities of the day being wet or dry is independent of the initial conditions of the weather.

Table 2.1 Year wise transition counts for the Stations under study

Station	Imphal				Mohanbari				Guwahati				Cherrapunji			
Year	n_{00}	n_{01}	n_{10}	n_{11}	n_{00}	n_{01}	n_{10}	n_{11}	n_{00}	n_{01}	n_{10}	n_{11}	n_{00}	n_{01}	n_{10}	n_{11}
2001	47	25	25	25	37	21	20	44	54	23	24	21	8	12	12	90
2002	50	21	21	30	40	24	23	35	51	25	25	21	14	15	14	79
2003	40	23	23	36	21	25	24	52	43	25	25	29	11	13	12	86
2004	49	27	27	19	35	22	23	42	60	21	22	19	9	16	16	81
2005	50	25	25	22	43	18	18	43	51	21	21	29	24	12	12	74

Table 2.2 Year wise transition probabilities for the stations under study

Station	Imphal				Mohanbari			
Year	$1-p_0$	p_0	$1-p_1$	p_1	$1-p_0$	p_0	$1-p_1$	p_1
2001	.65	.35	.50	.50	.64	.36	.31	.69
2002	.70	.30	.41	.59	.62	.38	.40	.60
2003	.63	.37	.39	.61	.46	.54	.32	.68
2004	.64	.36	.59	.41	.61	.39	.35	.65
2005	.67	.33	.53	.47	.70	.30	.30	.70
Station	Guwahati				Cherrapunji			
Year	$1-p_0$	p_0	$1-p_1$	p_1	$1-p_0$	p_0	$1-p_1$	p_1
2001	.70	.30	.53	.47	.40	.60	.12	.88
2002	.67	.33	.54	.46	.48	.52	.15	.85
2003	.63	.37	.46	.54	.46	.54	.12	.88
2004	.74	.26	.54	.46	.36	.64	.16	.84
2005	.71	.29	.42	.58	.67	.33	.14	.86

Table 2.3 Calculated values of χ^2 for parameters p_0 and p_1

Sl. No.	Station	$\chi^2_{p_0}$	$\chi^2_{p_1}$	Tabulated χ^2 at 5% level of significance with 1 d.f.
1	Imphal	.69	29.12	.00393
2	Mohanbari	.75	.26	.00393
3	Guwahati	4.48	33.50	.00393
4	Cherrapunji	12.65	71.94	.00393

Table 2.4 Estimated parameters and different properties of 1st order two state Markov Chain

Station	p_0	p_1	Expected length of			Expected no. of		Actual no. of wet days
			Wet Spells	Dry Spells	Cycle	Dry days	Wet days	
Imphal	.34	.52	2.08	2.94	5.02	71.41	50.59	51
Mohanbari	.38	.67	3.03	2.63	5.66	56.7	65.30	65
Guwahati	.31	.50	2.00	3.23	5.23	75.31	46.69	47
Cherrapunji	.51	.86	7.14	1.96	9.10	26.28	95.72	96

Station	Standerd deviation of dry or wet days	Equilibrium State Probability		No.of days to Equilibrium
		π_0	π_1	
Imphal	6.63	.59	.41	7
Mohanbari	7.43	.46	.54	8
Guwahati	6.51	.62	.38	6
Cherrapunji	6.54	.22	.78	9

2.4 Conclusion

First order two state Markov chain Model is used to study the occurences of rainfall in Imphal, Mohanbari, Guwahati and Cherrapunji. For the study, five years (2001-2005) daily rainfall data in mm. over these four stations were collected from Regional Meteorological centre, Guwahati. Each day is classified as wet day if amount of rainfall is greater or equal to 3 mm and dry day if it is less than 3 mm. A sequence of wet and dry days for each stations over each year during the summer monsoon period (June-Sept.) was obtained and then using relative frequencies from the data over years, the probability p_0 of a wet day following a dry day and p_1 of a wet day following a wet day were calculated for each station separately. Then normal deviate test is applied to judge the efficiency of Markov chain in studying the pattern of occurrence of rainfall. The test result shows that occurrences of wet and dry days follow a two state Markov chain. Chi square tests were performed for the common estimates of the two parameters p_0 and p_1. But these tests indicated that none of the stations have similar patterns of the occurrences of rainfall. So they could not be pooled. The values of p_0 varies from

.31 to .38 for Imphal, Mohanbari, Guwahati and its value was .51 for Cherrapunji. The values of p_1 varies from .50 to .67 for the first three stations .86 for Cherrapunji. The expected length of dry spells is seen varying from 2.23 to 2.94 and the expected length of wet spells varies from 2.00 to 3.03 for the stations Imphal,Mohanbari and Guwahati and for Cherrapunji the values are 1.96 and 7.14 respectively. The expected numbers of wet day during the period of 122 days are calculated and compared with the actual values and found that the values are almost same. The number of wet days varies from 47 to 65 for Imphal, Mohanbari and Guwahati and 96 for Cherrapunji. The numbers of days to equilibrium are also calculated and found that their values vary between 6 to 9 for all the stations.

Chapter 3

Determination of the Order of a Markov Chain for Daily Rainfall Data: Application of Akaike Information Criterion

This chapter aims at demonstrating the application of the Akaike information criterion (AIC) to determine the order of two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons in North-East India. For the majority of the stations Markov chain of order one is identified as the most appropriate model, followed by order two, for describing the daily precipitations occurrences over North East India during Indian summer monsoon season.

3.1 Introduction

It is well known fact that Markov chain model can be fitted to daily rainfall occurrence and several authors used Markov chain model to estimate the wet and dry days in past, we refer to [9], [22], [23], [28] for first order Markov chain model and [37], [77] for higher order Markov chain model. Although a good number of literatures are available describing the Markov chain model for daily precipitation round the globe, no rigorous work barring the works by Bhargava et al. ([9]) and Medhi ([45]) pursued in India. The present study is an effort to demonstrate the application of higher order two state Markov chain over a series of daily rainfall data of seven stations in North East India.

In the previous chapter, we have seen that Markov chain model of order one can be fitted to the daily rainfall data over North East India. But it does not guarantee that we can ignore other higher order Markov chain model. Therefore an analysis regarding the fitting of Markov chain of appropriate order has been made in this chapter. We apply Markov chain of order up to three to the sequences of wet and dry days observed at seven distantly located stations in North East region of India. The best fitting model is then determined using the AIC by checking the minimum of AIC estimate and it is found that Markov chain of order one is an adequate model for most of the stations of North East regions of India to determine the daily precipitation.

A brief outline of this chapter is as follows. Section 3.2 introduces a brief specification of data set and the statistical methods used in this chapter. Section 3.3 is concerned with a discussion on the results obtained using the AIC criterion for different orders of Markov chain.

3.2 Data and Methodology

In this chapter, a series of daily rainfall data of seven stations in North East India viz. Imphal (2001-2005), Mohanbari (1993-2006), Guwahati (2001-2005), Cherrapunji (2001-2005), Silcoorie (1986-2005), North Bank (1986-2005), Tocklai (1986-2005) have been selected. The locations of these seven stations of North East India are shown in Figure 1.1. The series of daily rainfall are taken from Regional Meteorological Centre, Guwahati and Tocklai Experimental Station, Jorhat involving the aforesaid seven stations for the summer season (April to September) in each year. The Akaike information Criterion was introduced by Akaike [2] as an extension to final prediction error and since then it has been used successfully in various fields of statistics, engineering, hydrology and numerical analysis (cf. [2], [3], [55], [56], [62], [63]). The procedure for the determination of the order of a Markov chain by Akaike's Information Criterion (AIC) was developed by Tong ([71]). In the present work, application of the Akaike information criterion is demonstrated to determine the order of two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons in North-East India.

In statistical inference situations, Akaike ([2]) proposed the use of the entropy $B[\bullet]$ given by:

$$B[f;z] = \int \log\left\{\frac{g(z;x)}{f(z)}\right\} f(z)dz \tag{3.2.1}$$

where x is the vector of observations, and $f(z)$ and $g(z;x)$ are the probability density functions of the true and fitted models, respectively. According to the entropy maximization principle ([2]), the objects of statistical inference are to estimate $f(z)$ from the data x and to try to find $g(z;x)$ which maximizes the expected entropy:

$$E\{B[f;g]\} = \int B[f;g]f(x)dx \tag{3.2.2}$$

where E denotes the expectation operator and x is the vector of observations. Akaike ([2]) showed that for the number of observations $n \geq 30$:

$$-2nE\{B[f;g]\} \approx \eta + 2k - L, \tag{3.2.3}$$

η is a log-likelihood ratio test function given by:

$$\eta = -2\sum_{i=1}^{n} \log\left\{\frac{g(x_i|_k\hat{\theta})}{g(x_i|_L\hat{\theta})}\right\} \tag{3.2.4}$$

with

L = number of parameters (dimension) of the true model,

k = number of parameters (dimension) of the fitted model,

${}_k\hat{\theta}$ and ${}_L\hat{\theta}$ are the estimated parameters of the fitted and true models respectively.

Thus, from equation (3.2.3) and by ignoring the constant terms, Akaike derived a criterion which is now called the Akaike Information Criterion (AIC) given by:

$$AIC = -2\sum_{i=1}^{n}\log\{g(x_i|_k\hat{\theta})\} + 2k$$

which can also be written as:

$$\begin{aligned} \text{AIC} &= -2 \times \log(\text{Maximum likelihood for model}) \\ &\quad +2 \times (\text{Number of independent parameters in the model}). \qquad (3.2.5) \end{aligned}$$

This statistics was introduced as a measure of the deviation of the fitting model from the true structure. The first term on the right hand side of the equation (3.2.5) is a measure of the lack-of-fit of the chosen model, while the second term measures the increased unreliability of the chosen model due to the increased number of model parameters. The best approximating model is the one which achieves the minimum AIC in the class of the competing models. The procedure which, given several models, adopts the model that minimizes the AIC is called the minimum AIC estimate (MAICE). It is important to note that, since the AIC test is based on the maximum likelihood function, which is asymptotically effective and unbiased, the test yields fairly accurate results for $n \geq 30$, where n is the number of observations. However, the test has been used with considerable success for $n \geq 20$ (cf. [39]).

Denote the transition probability for a r order chain by $p_{ij\ldots kl}$, $i = 1, 2, \ldots, s$, s being the finite number of states of the chain and the suffix contains r+1 characters. Then the maximum likelihood estimates of $p_{ij\ldots kl}$ is given by

$$\hat{p}_{ij\ldots kl} = \frac{n_{ij\ldots kl}}{n_{ij\ldots k\cdot}},$$

where $n_{ij\ldots kl}$ is the number of transition from the state i to the states l through the state $j \ldots k$ and $n_{ij\ldots k\cdot} = \sum_l n_{ij\ldots kl}$. The hypothesis tested is $H_{r-1} : p_{ij\ldots kl} = p_{j\ldots kl}$, $i = 1, 2, \ldots, s$ (that the chain (r-1) dependent against H_r: that the chain is r dependent). The statistics constructed is

$${}_{r-1}\eta_r = -2\log\lambda_{r-1,r} = 2\sum_{i,j,\ldots,k,l} n_{ij\ldots kl}\Big(\log\frac{n_{ij\ldots kl}}{n_{ij\ldots k\cdot}} - \log\frac{n_{\cdot j\ldots kl}}{n_{\cdot j\ldots k\cdot}}\Big)$$

which is a χ^2 variable with $s^{r-1}(s-1)^2$ degrees of freedom under H_{r-1}. The hypothesis H_k that the chain is k dependent implies the hypothesis H_r that the chain is r dependent, whenever $k < r$. Hence, the hypothesis H_k is a subset of the hypothesis H_r. Denote by $\lambda_{k,r}$, the ratio of the maximum likelihood given H_k to that given H_r, then

$$\begin{aligned} \lambda_{k,r} &= \lambda_{k,k+1}\lambda_{k+1,k+2}\ldots\lambda_{r-1,r} \quad \text{and so} \\ {}_k\eta_r &= -2\log\lambda_{k,k+1} - 2\log\lambda_{k+1,k+2} - \ldots - 2\log\lambda_{r-1,r}, \quad k < r. \end{aligned}$$

Again, under H_k, Good ([27]) has shown that $-2\log\lambda_{k,r}$ i.e., ${}_k\eta_r$ has a χ^2 variable with degrees of freedom

$$\begin{aligned} &\nabla s^{r+1} - \nabla s^{k+1} \quad \text{for } k \geq 0 \\ \text{and} \quad &\nabla s^{r+1} \qquad\qquad\quad \text{for } k = -1, \end{aligned}$$

under H_k, where ∇ is the standard backward operator given by $\nabla s^r = s^r - s^{r-1}$.

If the statistical identification procedure is considered as a decision procedure, the most basic problem is the appropriate choice of the risk (expected loss) function. The loss functions considered in classical theory of hypothesis testing are defined by the probabilities of accepting the incorrect hypothesis or rejecting the correct hypothesis.

Tong ([71]) proposes the choice of the loss function, based on AIC approach as

$$R(k) =_k \eta_M - 2(\nabla S^{M+1} - \nabla S^{k+1}),$$

where M is the highest order model to be considered and k is the order of the fitting model. The minimum AIC estimate (MAICE) of the order of the Markov chain is that value of k which gives the minimum of $R(k)$ over all orders considered. Raising the order of Markov chain does not necessarily do away the imperfections of the model. On the other hand, the number of parameters to estimate increases with 2^k for two state, k order Markov chain which may rapidly enhance the uncertainty of the estimation. Therefore the present study is confined to the Markov chain of order up to three.

Table 3.1 Likelihood statistic for North Bank

Year	$_0\eta_1$	$_0\eta_2$	$_0\eta_3$	$_1\eta_2$	$_1\eta_3$	$_2\eta_3$	$_3\eta_3$
1986	16 7276	16 8175	19 8498	0 0899	3 1222	3 0323	0
1987	17 7543	22 8454	28 7556	5 0911	11 0013	5 9102	0
1988	41 1159	42 1816	50 9963	1 0657	9 8804	8 8147	0
1989	19 1132	24 3546	32 6416	5 2414	13 5284	8 287	0
1990	21 1622	21 2464	27 8429	0 0842	6 6807	6 5965	0
1991	6 2446	12 6413	21 1247	6 3967	14 8801	8 4834	0
1992	4 6952	9 8015	11.2007	5 1063	6 5055	1 3992	0
1993	18 9827	23 6082	25 9841	4 6255	7 0014	2 3759	0
1994	20 5584	23 2673	26 4903	2 7089	5 9319	3 223	0
1995	28 3186	36 5763	44 1681	8 2577	15 8495	7 5918	0
1996	23 5193	24 5876	27 8354	1 0683	4 3161	3 2478	0
1997	17 4515	19 047	26 7143	1 5955	9 2628	7 6673	0
1998	42 0208	45 0825	50 9531	3 0617	8 9323	5 8706	0
1999	12 1565	16 5651	20 6124	4 4086	8 4559	4 0473	0
2000	4 8409	8 9134	15 9362	4 0725	11 0953	7 0228	0
2001	6 8661	7.3759	9.4704	0 5098	2 6043	2 0945	0
2002	22 3583	22 542	26 9619	0 1837	4 6036	4 4199	0
2003	17 034	19 485	23 7815	2 451	6 7475	4 2965	0
2004	25 857	29 9857	30 2572	4 1287	4 4002	0 2715	0
2005	12 9343	16 5079	20 4342	3 5736	7 4999	3 9263	0
df	1	3	7	2	6	4	0

Table 3.2 AIC values for the station Cherrapunji

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
2001	25.296	-5.5744	-4.7736	0	-5.5744	1
2002	16 5769	1.8441	1.4442	0	0	3
2003	16.7548	-0.0637	-1.2467	0	-1.2467	2
2004	0 7354	-4.4915	-0.9573	0	-4.4915	1
2005	21.1522	-2.5002	-3.3747	0	-3.3747	2

Table 3.3 AIC values for the station Guwahati

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
2001	-1.0056	-5 0543	-4 2653	0	-5.0543	1
2002	10.4165	-4 4754	-1 8223	0	-4.4754	1
2003	17 846	11.4287	13.3948	0	0	3
2004	-1.36577	-5.60417	-1.8506	0	-5.60417	1
2005	18.7865	7.188	2.8948	0	0	3

Table 3.4 AIC values for the station Imphal

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
2001	4 2847	-6 3741	-2 5225	0	-6 3741	1
2002	32 9413	8 7132	-5 0412	0	-5 0412	2
2003	16 9987	2 6758	0 1431	0	0	3
2004	9 5471	1 3503	2 0861	0	0	3
2005	12 7122	-4 5871	-2 0123	0	-4 5871	1

Table 3.5 AIC values for the station Mohanbari

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
1993	6 3891	-3 1913	-3 0882	0	-3 1913	1
1994	5 4926	-6 3	-5 2347	0	-6 3	1
1995	7 322	-1 9796	-0 0287	0	-1 9796	1
1996	13 724	-7 2275	-6 3125	0	-7 2275	1
1997	7 6405	-5 1164	-1 9863	0	-5 1164	1
1999	5 5773	2 508	-4 5951	0	-4 5951	2
2001	-2 6165	-10 0504	-7 9391	0	-10 0504	1
2002	7 6799	-10 6054	-6 9666	0	-10 6054	1
2003	8 2287	-0 0316	-5 7523	0	-5 7523	2
2004	15 5802	0 3969	1 9405	0	0	3
2005	7 4747	1 6295	4 336	0	0	3
2006	17 0646	-0 4816	-0 8071	0	-0 8071	2

Table 3.6 AIC values for the station Northbank

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
1986	5 8498	-8 8778	-4 9677	0	-8 8778	1
1987	14 7556	-0 9987	-2 0898	0	-2 0898	2
1988	36 9963	-2 1196	0 8147	0	-2 1196	1
1989	18 6416	1 5284	0 287	0	0	3
1990	13 8429	-5 3193	-1 4035	0	-5 3193	1
1991	7 1247	2 8801	0 4834	0	0	3
1992	-2 7993	-5 4945	-6 6008	0	-6 6008	2
1993	11 9841	-4 9986	-5 6241	0	-5 6241	2
1994	12 4903	-6 0681	-4 777	0	-6 0681	1
1995	30 1681	3 8495	-0 4082	0	-0 4082	2
1996	13 8354	-7 6839	-4 7522	0	-7 6839	1
1997	12 7143	-2 7372	-0 3327	0	-2 7372	1
1998	36 9531	-3 0677	-2 1294	0	-3 0677	1
1999	6 6124	-3 5441	-3 9527	0	-3 9527	2
2000	1 9362	-0 9047	-0 9772	0	-0 9772	2
2001	-4 5296	-9 3957	-5 9055	0	-9 3957	1
2002	12 9619	-7 3964	-3 5801	0	-7 3964	1
2003	9 7815	-5 2525	-3 7035	0	-5 2525	1
2004	16 2572	-7 5998	-7 7285	0	-7 7285	1
2005	6 4342	-4 5001	-4 0737	0	-4 5001	1

Table 3.7 AIC values for the station Silcoorie

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
1986	6 1958	-5 4116	-5 222	0	-5 4116	1
1987	12 6532	-4 9303	-5 5398	0	-5 5398	2
1988	3 3324	-7 0767	-4 8481	0	-7 0767	1
1989	14 4774	-5 8251	-5 4369	0	-5 8251	1
1990	5 8544	-8 6228	-4 7136	0	-8 6228	1
1991	0 4625	-9 8248	-7 3887	0	-9 8248	2
1992	13 3544	0 621	-7 6684	0	-7 6684	2
1993	13 737	0 3297	-3 6292	0	-3 6292	2
1994	14 1871	-4 5688	-2 558	0	-4 5688	1
1995	-1 9577	-9 4869	-6 0686	0	-9 4869	1
1996	27 6852	-1 9304	-5 0619	0	-5 0619	2
1997	5 6338	-5 7394	-4 8588	0	-5 7394	1
1999	24 1983	-3 3469	-2 7311	0	-3 3469	1
2001	7 5898	-0 8134	-4 293	0	-4 293	2
2002	11 6622	-5 0198	-7 2194	0	-7 2194	2
2003	16 485	-0 6414	-3 548	0	-3 548	2
2004	3 6647	-0 6573	1 2379	0	-0 6573	1
2005	44 7494	6 1562	9 6934	0	0	3

Table 3.8 AIC values for the station Tocklai

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
1986	-2 6125	-5 3556	-4 1269	0	-5 3556	1
1987	1 5623	-6 9284	-6 0647	0	-6 9284	1
1988	6 4774	-7 2628	-4 5987	0	-7 2628	1
1989	0 8393	-10 9265	-7 5985	0	-10 9265	1
1990	10 4431	6 0319	3 7292	0	0	3
1991	-2 322	-3 1392	0 179	0	-3 1392	1
1992	-4 6171	-5 4782	-3 4837	0	-5 4782	1
1993	-0 3524	-2 3872	-4 0922	0	-4 0922	2
1994	2 2017	-2 4974	-3 6537	0	-3 6537	2
1995	1 1907	-6 3817	-5 8844	0	-6 3817	1
1996	8 1645	1 0601	-4 7042	0	-4 7042	2
1997	1 08526	-5 95314	-3 40004	0	-5 95314	1
1998	4 8008	3 4529	0 8996	0	0	3
1999	5 4214	2 0093	-3 656	0	-3 656	2
2000	2 1448	-0 7117	1 128	0	-0 7117	1
2001	-5 6387	-5 6057	-6 1419	0	-6 1419	2
2002	3 7916	-4 6991	-1 9707	0	-4 6991	1
2003	-0 3613	-7 6114	-4 3962	0	-7 6114	1
2004	-5 4475	-6 7144	-7 4141	0	-7 4141	2
2005	0 2872	-2 3047	-5 778	0	-5 778	2

Table 3.9 Percentages of the best fitting orders of Markov Chain

Order of MC	North Bank	Tocklai	Silcoorie	Mohanbari	Cherrapunji	Guwahati	Imphal
0	0	0	0	0	0	0	0
1	60	55	50	58	40	60	40
2	30	35	44	25	40	0	20
3	10	10	6	17	20	40	40

3.3 Results and Discussions

The present work involves the year wise estimation of the likelihood ratio statistic for first, second and third order two state Markov chain for each station. Table 3.1 illustrates the estimates of likelihood ratio statistic for station North Bank. It is interesting to see that the columns corresponding to ${}_0\eta_1$, ${}_0\eta_2$, ${}_0\eta_3$ are significant at 5% level of significance except for the years 1992 and 2001. For the simplicity of the exposition other tables are not included. The details can be found in Appendix. Therefore, we may note that the chain is at least of order one. This observation enters into the findings of the previous chapter. Then $R(k)$ values for each station over each year are calculated. The calculated values are displayed in Tables 3.2-3.8. Then according to MAICE procedure we adopt as the order that value of k which gives minimum $R(k)$ and those values for k are illustrated in column 7 of the Tables 3.2-3.8. Finally, the performance of the best fitting order of the Markov chain is given in Table 3.9. The present study leads to the following observations:

- Markov chain model can be fitted to daily rainfall occurrence of North East regions of India.
- The first order Markov chain model that has been used extensively, is an adequate model for most of the stations of North East regions of India to determine the daily precipitation.

Chapter 4

Use of Probability Distributions for the Analysis of Daily Rainfall Data

Daily rainfall data can be characterized by a probability distribution function known from the statistical literature. In this chapter, some well known probability distributions are considered to find the best fitting probability distribution function of the daily rainfall data. The gamma and Weibull distributions are observed to be competing each other and both are very close to the observed distributions as evinced by the graphical plots.

4.1 Introduction

In the previous two chapters, we have discussed the application of the two state Markov chain model to estimate the wet and dry days of seven distantly located stations in North East India. In this chapter, an attempt has been made to examine the goodness of fit of the distributions based on daily rainfall observations sampled from seven distantly located stations in North East Region of India Viz.Imphal, Mohanbari, Guwahati, Cherrapunji, Silcoorie, North Bank, Tocklai (Jorhat). Two-parameter gamma distribution, the left-truncated normal distribution, 2-parameter Weibull distribution and 2-parameter lognormal distribution is considered to find the best fitting probability distribution function of the daily rainfall data. Chi-square test and Kolmogorov-Smirnov test have been performed judging the goodness of fit. Cumulative distribution functions (cdf) for each of the aforesaid distributions and the observed cumulative distribution functions are plotted for identifying the right probability density function for the daily rainfall amount. For the literature concerning the fitting of probability distribution of daily rainfall, we refer to [5], [6], [49], [51], [65], [67] and [69]. So far no rigorous work barring the work by Medhi [45] pursued in the North East region of India, considerable effort has been made to graduate the rainfall of different time scales by fitting an appropriate frequency distributions. The two basic objectives of this chapter are to judge the goodness of fit of the distributions fitted for daily rainfall observations sampled from seven stations of North East India and to detect the competing distributions.

The organization of this chapter is as follows: In Section 4.2, we introduce the data sets, probability distributions and related probability density functions, goodness of fit of the distributions used in this chapter. While Section 4.3 is concerned with the findings of goodness of fit test, the chapter ends with a concluding remark in Section 4.4.

4.2 Data and Methodology

In this study, seven distantly located stations in North East India viz. Imphal, Mohanbari, Guwahati, Cherrapunji, Silcoorie, North Bank, Tocklai (Jorhat) have been selected. The locations of these seven stations of North East India are shown in Figure 1.1. The study utilizes daily rainfall data in mm for five years (2001-2005). The series of daily rainfall are taken from Regional Meteorological Centre, Guwahati and Tocklai Experimental Station, Jorhat involving the aforesaid seven stations for the summer monsoon months of June, July, August and September in each year.

4.2.1 Left-truncated Normal Distribution

The probability density function of a normally distributed random variable x is given by

$$y = f(x/\mu_n, \sigma_n) = \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu_n}{\sigma_n}\right)^2 \right\}, \; \sigma_n > 0, -\infty \leq x \leq \infty. \quad (4.2.1)$$

If the values of x below some value x_L cannot be observed due to truncation then, the resulting distribution is a left-truncated normal distribution with probability density function f_{LTN} given by (4.2.2)

$$f_{LTN}(x) = \begin{cases} 0 & -\infty \leq x \leq x_L \\ f(x)/\int_{x_L}^{\infty} f(x)dx & x_L \leq x \leq \infty, \end{cases} \quad (4.2.2)$$

where $f(x)$ is as defined in Equation (4.2.1), and μ_n and σ_n are the parameters of the distribution and are equal to mean $\bar{X}$ and standard deviation $\sqrt{V(X)}$ of the sampling distribution respectively.

4.2.2 Lognormal Distribution

The probability density function of the lognormal distribution is given by

$$y = f(x/\mu_l, \sigma_l) = \frac{1}{x\sigma_l\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{z-\mu_l}{\sigma_l}\right)^2 \right\}, \quad (4.2.3)$$

where $z = \log x$, and μ_l and σ_l are the parameters of the distribution and can be evaluated using the following relationship

$$\hat{\mu}_l = \bar{z} \quad (4.2.4)$$

$$\hat{\sigma}_l = \left[n^{-1} \sum_{j=1}^{n} (z_j - \bar{z})^2 \right]^{1/2} \quad (4.2.5)$$

where $\bar{z} = n^{-1} \sum_{j=1}^{n} z_j$ (assuming that $x_1, x_2, \ldots, x_n$ are independent random variables each having the same lognormal distribution).

4.2.3 Gamma Distribution

Gamma distribution is next to the normal distribution in simplicity and the same time it covers a wide range of skewness. We therefore decided to test the fit of daily rainfall to gamma distribution for which the probability density function is given by

$$\begin{aligned} f(x/\lambda,\eta) &= \frac{1}{\eta^{\lambda}\Gamma(\lambda)}x^{\lambda-1}e^{-x/n}, \quad x,\lambda,\eta>0 && (4.2.6)\\ &= 0 \qquad\qquad \text{otherwise} && (4.2.7)\end{aligned}$$

where η and λ are scale and shape parameters, respectively. The exponential distribution is a particular case when $\lambda = 1$. The maximum likelihood estimates $\hat{\lambda}$ and $\hat{\eta}$ of the parameters can be obtained by solving the equations

$$\begin{aligned} n^{-1}\sum_{j=1}^{n}\log X_j &= \log\hat{\eta} + \Psi(\hat{\lambda}), && (4.2.8)\\ \bar{X} &= \hat{\lambda}\hat{\eta} && (4.2.9)\end{aligned}$$

where $\bar{X}$ is the arithmetic mean of the rainfall amounts $x_1, x_2, \ldots, x_n$ and

$$\Psi(\hat{\lambda}) = \frac{\partial\log\Gamma(\hat{\lambda})}{\partial\hat{\lambda}} = \text{di-gamma function.} \qquad (4.2.10)$$

4.2.4 Weibull Distribution

The probability density function of the two parameter Weibull distribution is given by

$$y = f(x/\alpha,\beta) = \alpha\beta x^{\beta-1}\exp(-\alpha x^{\beta}), \quad x>0$$

where α be the scale parameter and β be the shape parameter of the distribution. The maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β respectively satisfy the equations

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^{n} x_i^{\hat{\beta}}\log x_i}{\sum_{i=1}^{n} x_i^{\hat{\beta}}} - \frac{\sum_{i=1}^{n}\log x_i}{n}\\ \hat{\alpha} &= \frac{n}{\sum_{i=1}^{n} x_i^{\hat{\beta}}}.\end{aligned}$$

Table 4.1 Fitting of probability distributions for daily rainfall data (2001-2005) of Mohanbari during Indian Summer Monsoon Season.

Rainfall (mm.)	Observed. frequencies	Theoretical frequencies.			
		Truncated Normal. μ_n =16.6879 σ_n =21.2957	Lognormal μ_l=1.9568 σ_l=1.5023	Gamma λ =.7039 η =23.7079	Weibull α =.1224 β =.7859
0-14	276	132	300	269	277
14-28	87	143	64	90	85
28-42	34	102	28	42	39
42-56	21	48	15	21	20
56-70	11	15	9	11	10
70-84	4	3	6	6	6
84-98	8	0	4	3	3
98-112	2	0	3	2	2
112-126	0	0	2	1	1
126-140	2	2	14	0	2
Kolmogorov -Smirnov D Statistics		.3226	.0549	.0164	.0086
χ^2		264.8085	25.3793	8.4726	3.5084
d.f		3	5	4	4
p-value		4.09828e-57	0.00012	0.07572	0.47660

Table 4.2 Fitting of probability distributions for daily rainfall data (2001-2005) of Guwahati during Indian Summer Monsoon Season.

Rainfall (mm.)	Observed. frequencies	Theoretical frequencies.			
		Truncated Normal μ_n =13.0294 σ_n =19.6634	Lognormal μ_l=1.5200 σ_l=1.6097	Gamma. λ =.5913 η =22.0360	Weibull α =.2012 β =.6952
0-14	282	141	300	273	284
14-28	59	136	45	68	61
28-42	25	81	18	29	25
42-56	12	30	10	13	12
56-70	9	7	6	6	6
70-84	5	1	4	3	3
84-98	0	0	3	2	2
98-112	1	0	2	1	1
112-126	1	0	1	0	1
126-140	2	0	7	1	1
Kolmogorov -Smirnov D Statistics		.3558	.0444	.0235	.0089
χ^2		246.6116	14.5292	4.1876	1.7047
d.f		2	4	3	3
p-value		2.81172e-54	0.00578	0.24191	0.63589

Table 4.3 Fitting of probability distributions for daily rainfall data (2001-2005) of Imphal during Indian Summer Monsoon Season.

Rainfall (mm.)	Observed. frequencies	Theoretical frequencies.			
		Truncated Normal μ_n =10.5100 σ_n =16.3932	Lognormal μ_l=1.3093 σ_l=1.6168	Gamma λ =.5934 η =17.7122	Weibull α =.2321 β =.6976
0-14	329	189	342	321	331
14-28	61	159	43	68	60
28-42	21	67	17	25	22
42-56	9	14	9	10	9
56-70	4	2	5	4	4
70-84	3	0	3	2	2
84-98	1	0	2	1	1
98-112	2	0	2	0	1
112-126	0	0	1	0	0
126-140	1	0	7	0	1
Kolmogorov –Smirnov D Statistics		.3257	.0312	.0177	.0051
χ^2		196.6883	14.8131	3.9457	.5187
d.f		1	4	2	2
p-value		1.10293e-44	0.00510	0.13906	0.77155

Table 4.4 Fitting of probability distributions for daily rainfall data (2001-2005) of Cherrapunji during Indian Summer Monsoon Season.

Rainfall (mm.)	Observed. frequencies	Theoretical frequencies.			
		Truncated Normal μ_n =77.7161 σ_n =106.9166	Lognormal μ_l=3.3920 σ_l=1.6223	Gamma λ =.6373 η =121.9516	Weibull α =.0469 β =.7360
0-80	373	190	386	355	367
80-160	74	187	64	96	88
160-240	39	108	27	41	37
240-320	19	36	15	19	17
320-400	13	7	9	9	9
400-480	5	1	6	4	5
480-560	3	0	4	2	3
560-640	1	0	3	1	2
640-720	1	0	2	1	1
720-800	1	0	13	1	0
Kolmogorov –Smirnov D Statistics		.3464	.0337	.0334	.0151
χ^2		328.652	23.1995	8.2741	4.4465
d.f		2	5	3	4
p-value		4.30651e-72	0.00031	0.04067	0.34893

Table 4.5 Fitting of probability distributions for daily rainfall data (2001-2005) of Silcoorie during Indian Summer Monsoon Season.

Rainfall (mm.)	Observed. frequencies	Theoretical frequencies.			
		Truncated Normal. μ_n=15.5706 σ_n=18.0755	Lognormal μ_l=2.1468 σ_l=1.1634	Gamma λ=.9681 η=16.0844	Weibull α=.0771 β=.9434
0-14	291	153	303	272	276
14-28	97	163	83	108	104
28-42	27	98	31	45	43
42-56	21	34	15	18	18
56-70	9	6	8	8	8
70-84	7	1	5	3	4
84-98	2	0	3	1	2
98-112	1	0	2	1	1
112-126	0	0	1	0	0
126-140	1	1	5	0	0
Kolmogorov –Smirnov D Statistics		.3019	.0257	.0423	.0328
χ^2		225.6039	11.3778	17.4726	10.1506
d.f		2	5	3	3
p-value		1.02503E-49	0.0444	0.0002	0.0173

Table 4.6 Fitting of probability distributions for daily rainfall data (2001-2005) of North Bank during Indian Summer Monsoon Season.

Rainfall (mm.)	Observed. frequencies	Theoretical frequencies.			
		Truncated Normal. μ_n=17.8168 σ_n=21.6805	Lognormal μ_l=2.0529 σ_l=1.4704	Gamma λ=.7268 η=24.5152	Weibull α=.1088 β=.8044
0-20	269	169	297	277	282
20-40	81	155	51	76	71
40-60	30	64	20	29	27
60-80	16	12	10	12	11
80-100	3	1	6	5	5
100-120	2	0	4	2	2
120-140	0	0	3	1	1
140-160	0	0	2	0	1
160-180	0	0	1	0	0
180-200	1	1	8	0	2
Kolmogorov –Smirnov D Statistics		.2481	.0702	.0199	.0331
χ^2		117.1346	42.9118	2.4278	6.9138
d.f		1	4	2	3
p-value		12.6824E-27	1.0793E-08	0.2970	0.0747

Table 4.7 Fitting of probability distributions for daily rainfall data (2001-2005) of Tocklai during Indian Summer Monsoon Season.

Rainfall (mm.)	Observed. frequencies	Theoretical frequencies.			
		Truncated Normal. μ_n=13.0146 σ_n=17.1076	Lognormal μ_l=1.6966 σ_l=1.4449	Gamma λ=.6957 η=18.7077	Weibull α=.1556 β=.7732
0-12	271	135	292	261	270
12-24	69	140	58	81	76
24-36	33	86	22	34	31
36-48	13	42	14	19	18
48-60	12	9	7	8	8
60-72	12	1	5	4	4
72-84	0	0	3	2	2
84-96	1	0	2	1	1
96-108	1	0	2	1	1
108-120	1	0	8	2	2
Kolmogorov -Smirnov D Statistics		.3305	.0511	.0249	.0206
χ^2		254.6012	32.2358	8.5851	6.6664
d.f		2	5	3	3
p-value		5.17681E-56	5.3355E-06	0.0353	0.0833

4.2.5 Test for goodness of fit

The tests applied for judging the goodness of fit of the distributions for rainfall series are namely chi-square test and Kolmogorov-Smirnov test. However, chi-square test has been carried out with caution considering its limitations in application and the suggestion made by Massey and Frank [44]. The authors of [44] showed that Kolmogorov-Smirnov test treats individual observation separately leading to no loss of information in grouping while loss of information in chi-square procedure is large. Recently, Pal [57] mentioned that the Chi square test's sensitivity to very small cell frequencies make itself unsuitable when expected frequencies work out at less than 5 in 20 percent of the cells. In the present case it is found that more than 50% of the cell frequencies are less than 5. Also according to Keeping [38], Kolmogorov Smirnov test can be applied in situations where the theoretical distribution function is continuous. Here also the theoretical distribution functions considered are continuous since the parameters are positive and x can assume values greater than zero. The test statistic used is

$$D_n = \max |S_n(x) - F(x)|$$

where $S_n(x)$ and $F(x)$ are empirical and theoretical distribution functions, respectively. The distribution of D_n is independent of $F(x)$. The theoretical distribution function however, has to be completely specified. In this study the theoretical distribution function have been calculated by using the estimated parameters of the distribution in each case. The significance of a critical value of D_n depends on n, the number of observations. If n is over 35, the critical values of D at .05 level of significance can be determined by the formula $1.36/\sqrt{n}$. Any D_n equal to or greater than $1.36/\sqrt{n}$ will be significant at .05 level (two tailed test).

4.3 Results

A day with rainfall of more than 0 mm or a trace be designated as a rainy day and with no rainfall as dry. After defining a rainy day, it is necessary to determine the amount of rainfall on such a day. In the present study, different distributions are considered as the probability distribution function of the daily rainfall data. The parameters for each distribution are estimated by maximum likelihood method from the daily rainfall data for each station separately and are provided in Tables 4.1 to 4.7. The tables also include the observed frequencies, expected frequencies obtained from the different fitted distributions. The values of Kolmogorov- Smirnov D-statistics, values of χ^2 along with degrees of freedom and the corresponding p-value for χ^2 are also provided as an evidence in support of goodness of fit. The Chi-square test of goodness of fit is applied to daily rainfall. The number of class intervals was found to be 10 over which the computations were done. Further, more than 50% of the cell frequencies were found to be less than 5 for almost every cases. Accordingly, Kolmorgorov-Smirnov test was applied in all cases following by the suggestion made by Pal [57]. Barring truncated normal distribution, other distributions viz. log-normal, gamma, and Weibull have been found satisfactory to model the rainfall series as evidenced by Kolmogorov-Smirnov test.

In order to confirm the goodness of fit for the above three distributions we additionally applied graphical plots of theoretical and observed cumulative distribution functions. The estimation of the cumulative distribution functions, $S_n(x) = P(X \leq x)$ for various preassigned values of x for each distribution viz. normal, log-normal, gamma, and Weibull distribution was calculated and graphs were drawn taking probabilities as ordinate and rainfall amount as abscissa (c.f. Figure 4.1). Graphic plots for probability density functions have been also shown in Figure 4.2. The computations have been carried out in the workstation Matlab 6.

The following salient features have been revealed from the goodness of fit tests and graphs:

- In general, truncated normal distribution appears to poorly represent the distribution of daily rainfall as evidenced by the tests and evinced by the graphs for pdf and cdf.

- The gamma and Weibull pdf can be regarded to compete with each other as both of them preserve the 'sigmoid' shape of the observed cumulative distribution of daily rainfall series. Also it is seen that gamma and Weibull pdf's are quite close to the observed pdf plots.

- Lognormal distribution, although accepted to be well fitted on the basis of chi-square and Kolmogorov-Smirnov test, does not seem to compete with gamma and Weibull distribution and also is observed to be quite a distance from the observed plot.

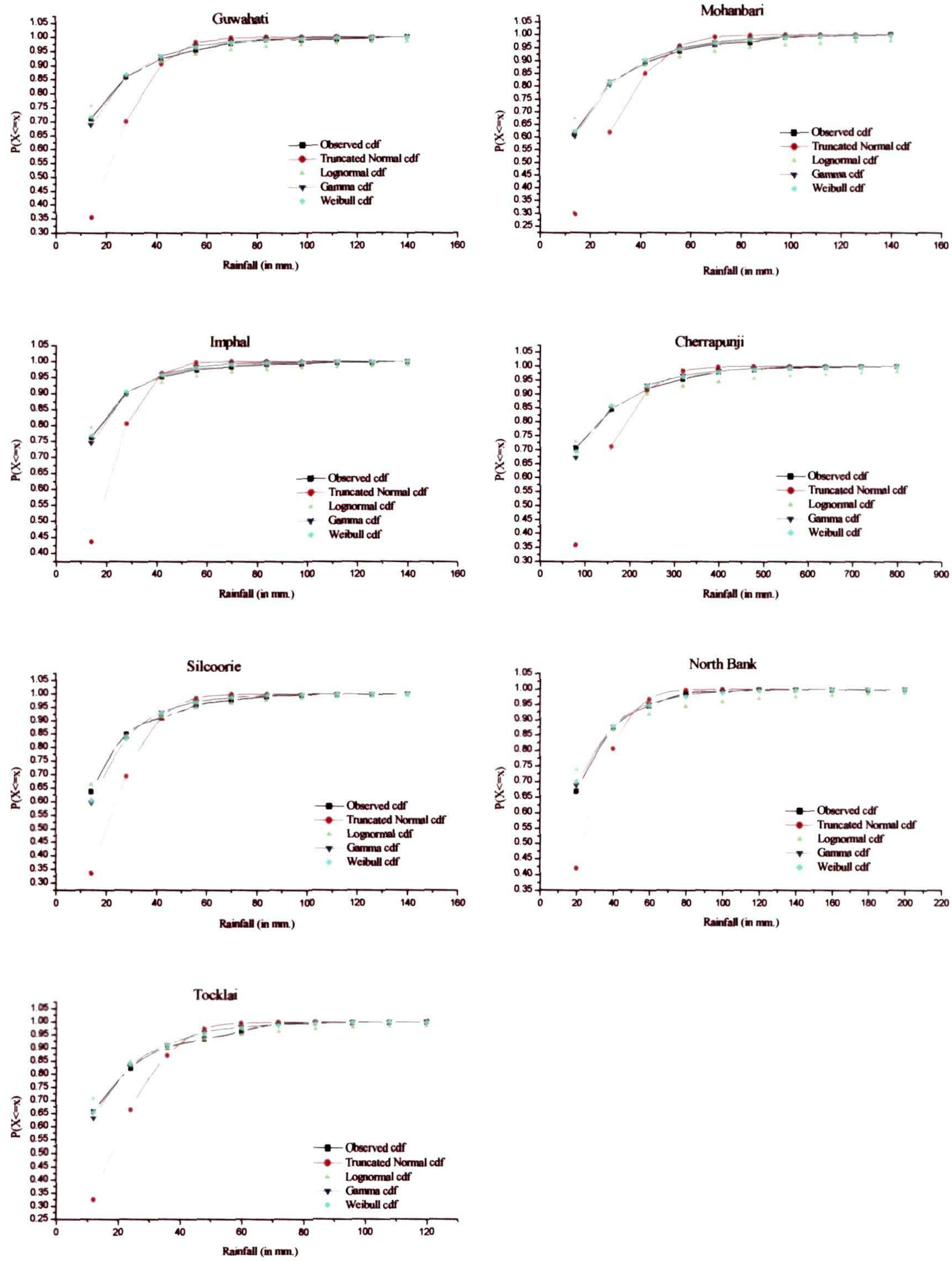

Figure 4.1 Curve for Probability Distribution Functions.

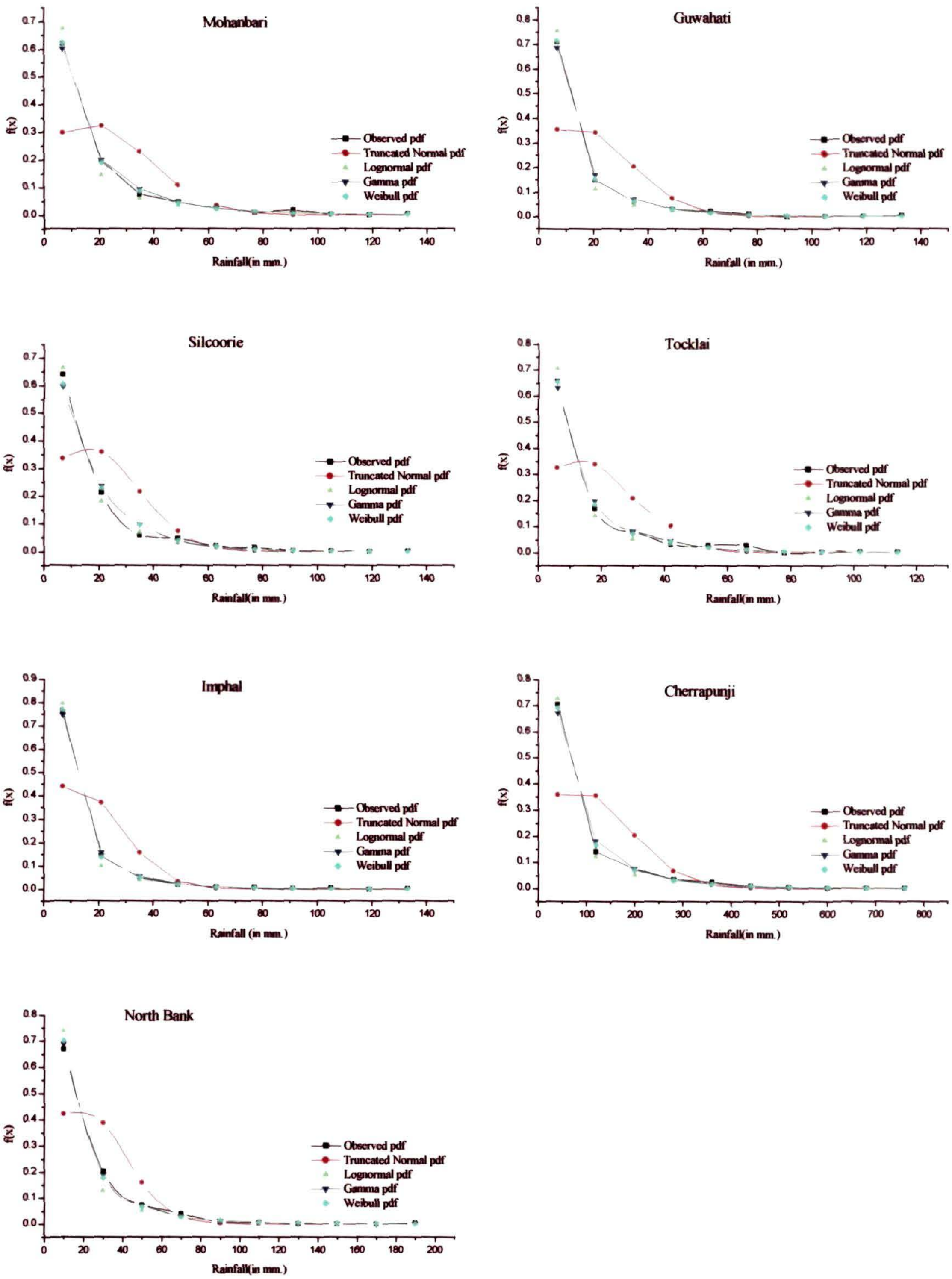

Figure 4.2 Curve for Probability Density Functions.

4.4 Conclusion

The following conclusions are drawn on the basis of the results and analysis made in this study.

The gamma and Weibull distributions are observed to be competing each other and both are very close to the observed distributions. It is well evidenced by the graphic plots animated on the basis of cdf and pdf. So far as goodness of fit of these two distributions are concerned, they are judged to be well fitted as evidenced by chi-square and Kolmogorov-Smirnov test.

Chapter 5

Statistical Modeling of Wet and Dry Spell Frequencies

In this chapter, an attempt has been made to develop a discrete precipitation model for the daily series of precipitation occurrences over North East India. Various distributions have been fitted to describe the wet and dry spell frequencies of occurrences. The goodness of fit of the proposed models have been tested using Kolmogorov-Smirnov test. It is observed that Eggenberger-Polya distribution fairly fits wet and dry spell frequencies and can be used in the future for an estimation of the wet and dry spells in the area under study.

5.1 Introduction

In the previous chapters, we have discussed the statistical modeling of daily precipitation occurrences over North East India. The point of approach in the present chapter is to model the duration of consecutive dry and wet days i.e. spell, instead of individual wet and dry days. The definition of spell is based on the duration of consecutive wet and dry days. A wet spell is a sequence of wet days and it begins and ends the day after and the day before a dry day. In this study a wet day (W) is considered as one where the precipitation is $\geq$ 1mm and, obviously, dry day (D) the one where there is no precipitation or is $\not\geq$ 1mm.

The main objective of the present study is to find the best fitting model to describe the wet and dry spell frequencies of occurrences considering the climatic features of the different parts of North-East India. Among the possible statistical models, the following models have been tested:

- Discrete uniform distribution
- Geometric distribution
- Logarithmic series
- Negative binomial distribution
- Poisson distribution
- Markov chain of order one and two
- Eggenberger-Polya distribution

The models are fitted to the observed data of seven stations namely Imphal, Mohanbari, Guwahati, Cherrapunji, Silcoorie, North Bank and Tocklai (Jorhat) of North-East India with pronounced attention to summer monsoon season. Then the Kolmogorov-Smirnov test for goodness of fit was employed as the significance test for every model, assuming the level of significance as 5% (α = .05). For the earlier literature concerning the statistical modeling of wet and dry spells, we refer to Section 1.4.

A brief outline of this chapter is as follows. Section 5.2 introduces a brief specification of data set and the statistical methods used in this work. In section 5.3, a discussion is carried out on the results obtained from different statistical models applied to analyze the wet and dry spell frequencies. Finally, section 5.4 is devoted to a critical assessment of the results obtained in section 5.3.

5.2 Data and Methodology

In this study series of daily rainfall data of seven stations in North East India viz. Imphal (2001-2005), Mohanbari (1993-2006), Guwahati (2001-2005), Cherrapunji (2001-2005), Silcoorie (1986-2005), North Bank (1986-2005), Tocklai (1986-2005) have been selected. The locations of these seven stations of North East India are shown in Figure 1.1. The series of daily rainfall are taken from Regional Meteorological Centre, Guwahati and Tocklai Experimental Station, Jorhat involving the aforesaid seven stations for the summer season (April to September) in each year.

When a spell overlaps a seasonal change (that is, it includes the 31st of march and 1st of April or 30th September and the 1st of October) it is considered in its whole up to its modality change even if it reaches the following season and we include it in the season in which it develops longer. The sample gives the observed frequency of wet/dry spell of i length (where i goes from 1 to the longest spell). The i length spell can be considered as a casual variable and its probability density can be calculated with theoretical models. The models that have been used to describe the empirical data are uniform, geometric, logarithmic, negative binomial, Poisson, defined by (5.2.1)-(5.2.5), respectively. Further, following the trend of Berger et al. [8] the spell frequencies have also been analyzed by Eggenberger-Polya distribution (cf. (5.2.6)) and Markov chain of order one and two defined by (5.2.9) and (5.2.11), respectively.

$$P(X=k)=\frac{1}{b-a}, \quad \text{for } b-a \text{ points} \tag{5.2.1}$$

$$P_1(X=k)=p_1(1-p_1)^{k-1}, \;\; 0<p_1<1 \tag{5.2.2}$$

$$P_2(X=k)=\frac{-\theta^k}{k\log(1-\theta)}, \;\; 0<\theta<1 \tag{5.2.3}$$

$$P_3(X=k)=\begin{pmatrix} n+k-1 \\ k \end{pmatrix}\frac{p^n}{1-p^n}(1-p)^k, \; 0<p<1, \; n>0 \tag{5.2.4}$$

$$P_4(X=k)=\frac{e^{-\lambda}\lambda^k}{(1-e^{-\lambda})k!}, \; \lambda>0, \tag{5.2.5}$$

where $k = 1,2,3\ldots$ is defined as the number of consecutive days of which a spell is composed.

The Eggemberger-Polya distribution is:

$$P_5(X=k) = \frac{d^k}{(1+d)^{h/d+k}} \frac{\Gamma(h/d+k)}{k!\Gamma(h/d)} \tag{5.2.6}$$

where Γ is the Gamma distribution. Again, it follows from the argument of Giuseppe et al. [26] that Eggemberger-Polya distribution maintains the following recursive relation

$$P_5(1) = \frac{1}{(1+d)^{m/d}} \tag{5.2.7}$$

$$P_5(k) = \frac{m+(k-2)d}{(k-1)(1+d)} P_5(k-1), \tag{5.2.8}$$

where $(m+1)$ is the mean length of a spell, d is given by $\sigma^2/m-1$, σ^2 being the variance of sequences length.

In the case of first order Markov chain the probability that a dry spell will last exactly n days is given by

$$Q_n = p_{00}^{n-1}.p_{01} = p_{00}^{n-1}(1-p_{00}) \quad \text{for } n \geq 1 \tag{5.2.9}$$

where p_{00} is the probability of a dry day following a dry day and p_{01} the probability of a rainy day following a rainy day. The two parameters p_{01} and p_{11} are required to be estimated for describing the Markov chain of order one. One can estimate these parameters according to the principle of maximum likelihood estimation. The maximum likelihood estimate of p_{ij}, $(i,j=0,1)$ is given by

$$p_{ij} = \frac{n_{ij}}{\sum_{j=0}^{1} n_{ij}} = \frac{n_{ij}}{n_i}, \tag{5.2.10}$$

n_{ij} is the number of direct transition from the state i to the state j.

In the second order Markov chain the probability Q_n is expressed as

$$Q_n = p_{100}.p_{000}^{n-2}.p_{001} \quad \text{for } n \geq 2 \tag{5.2.11}$$

$$Q_1 = p_{101} \tag{5.2.12}$$

and the maximum likelihood estimate of p_{ijk} $(i,j,k=0,1)$ is given by

$$p_{ijk} = \frac{n_{ijk}}{\sum_{k=0}^{1} n_{ijk}} = \frac{n_{ijk}}{n_{ij.}}, \tag{5.2.13}$$

n_{ijk} is the number of transition from the state i to the state k through j. The first order Markov chain only takes into account the state wet or dry of the day preceding a given one. In the same way, the second order considers the states of the two preceding days. Raising the order of Markov chain does not necessarily do away the imperfections of the model. As the number of parameters to estimate increases with 2^k for two state, k order Markov chain which may rapidly enhance the uncertainty of the estimation. Therefore the present study is confined to the Markov chain of order one and two.

The Kolmogorov-Smirnov test for goodness of fit is then employed as the significance test for each model which is one of the most powerful non parametric tests for differences between two cumulative frequency distributions of the observed and estimated ones. It is already mentioned that the Chi square test's sensitivity to very small cell frequencies make itself unsuitable when expected frequencies work out at less than 5 in 20 percent of the cells. In this study, we have also observed that more than 20% of the cell frequencies are less than 5 and therefore the Kolmogorov-Smirnov test is applied to test the goodness of fit. The test statistics used is $D_n' = \max|S_n(x) - F(x)|$ with $S_n(x)$ and $F(x)$ are empirical and theoretical distribution functions, respectively. Like earlier, the theoretical distribution function have been calculated by using the estimated parameters of the distribution in each case. In the second phase, the goodness of fit has been tested by Kolmogorov-Smirnov statistics and results are summarized in Table 5.1 to Table 5.7.

Table 5.1 Results of the Kolmogorov-Smirnov (K-S) Tests for North Bank (1986-2005)

Summer Wet Spells			Summer Dry Spells		
Serial No.	Distributions	K-S Statistic	Serial No.	Distributions	K-S Statistic
1	Discrete Uniform	0.3636	1	Discrete Uniform	0.3750
2	Geometric	0.4056	2	Geometric	0.4866
3	Logarithmic	0.4208	3	Logarithmic	0.4922
4	Neg. Binomial	0.5633	4	Neg. Binomial	0.5096
5	Poisson	0.2100	5	Poisson	0.2817
6	M.C of order one	0.0402	6	M.C of order one	0.0661
7	M.C of order two	0.0306	7	M.C of order two	0.0226
8	Eggenberger-Polya	0.0178	8	Eggenberger-Polya	0.0121
Critical value at $\alpha = .05$		0.0545	Critical value at $\alpha = .05$		0.0545

Table 5.2 Results of the Kolmogorov-Smirnov (K-S) Tests for Tocklai (1986-2005)

Summer Wet Spells			Summer Dry Spells		
Serial No.	Distributions	K-S Statistic	Serial No.	Distributions	K-S Statistic
1	Discrete Uniform	0.3333	1	Discrete Uniform	0.3333
2	Geometric	0.4439	2	Geometric	0.5385
3	Logarithmic	0.4526	3	Logarithmic	0.5484
4	Neg. Binomial	0.3313	4	Neg. Binomial	0.4692
5	Poisson	0.2095	5	Poisson	0.3749
6	M.C of order one	0.0235	6	M.C of order one	0.0730
7	M.C of order two	0.0152	7	M.C of order two	0.0096
8	Eggenberger-Polya	0.0178	8	Eggenberger-Polya	0.0186
Critical value at $\alpha = .05$		0.0505	Critical value at $\alpha = .05$		0.0504

Table 5.3 Results of the Kolmogorov-Smirnov (K-S) Tests for Silcoorie (1986-2005)

Summer Wet Spells			Summer Dry Spells		
Serial No.	Distributions	K-S Statistic	Serial No.	Distributions	K-S Statistic
1	Discrete Uniform	0.3333	1	Discrete Uniform	0.4285
2	Geometric	0.3499	2	Geometric	0.5256
3	Logarithmic	0.3795	3	Logarithmic	0.5334
4	Neg. Binomial	0.4249	4	Neg. Binomial	0.4704
5	Poisson	0.2567	5	Poisson	0.3513
6	M.C of order one	0.0618	6	M.C of order one	0.0763
7	M.C of order two	0.0232	7	M.C of order two	0.0215
8	Eggenberger-Polya	0.0176	8	Eggenberger-Polya	0.0190
Critical value at $\alpha = .05$		0.0597	Critical value at $\alpha = .05$		0.0601

Table 5.4 Results of the Kolmogorov-Smirnov (K-S) Tests for Mohanbari (1993-2006)

Summer Wet Spells			Summer Dry Spells		
Serial No.	Distributions	K-S Statistic	Serial No.	Distributions	K-S Statistic
1	Discrete Uniform	0.3333	1	Discrete Uniform	0.3571
2	Geometric	0.3951	2	Geometric	0.3922
3	Logarithmic	0.4126	3	Logarithmic	0.4104
4	Neg. Binomial	0.4439	4	Neg. Binomial	0.4343
5	Poisson	0.2237	5	Poisson	0.2292
6	M.C of order one	0.0574	6	M.C of order one	0.0368
7	M.C of order two	0.0287	7	M.C of order two	0.0105
8	Eggenberger-Polya	0.0321	8	Eggenberger-Polya	0.0325
Critical value at $\alpha = .05$		0.0695	Critical value at $\alpha = .05$		0.0694

Table 5.5 Results of the Kolmogorov-Smirnov (K-S) Tests for Cherrapunji (2001-2005)

Summer Wet Spells			Summer Dry Spells		
Serial No.	Distributions	K-S Statistic	Serial No.	Distributions	K-S Statistic
1	Discrete Uniform	0.2963	1	Discrete Uniform	0.4000
2	Geometric	0.2365	2	Geometric	0.5643
3	Logarithmic	0.3066	3	Logarithmic	0.5807
4	Neg. Binomial	0.2176	4	Neg. Binomial	0.4495
5	Poisson	0.3510	5	Poisson	0.4220
6	M.C of order one	0.0777	6	M.C of order one	0.0485
7	M.C of order two	0.0582	7	M.C of order two	0.0388
8	Eggenberger-Polya	0.0541	8	Eggenberger-Polya	0.0317
Critical value at $\alpha = .05$		0.1338	Critical value at $\alpha = .05$		0.1338

Table 5.6 Results of the Kolmogorov-Smirnov (K-S) Tests for Guwahati (2001-2005)

Summer Wet Spells			Summer Dry Spells		
Serial No.	Distributions	K-S Statistic	Serial No.	Distributions	K-S Statistic
1	Discrete Uniform	0.3750	1	Discrete Uniform	0.3333
2	Geometric	0.4558	2	Geometric	0.5019
3	Logarithmic	0.4631	3	Logarithmic	0.5077
4	Neg. Binomial	0.5086	4	Neg. Binomial	0.4213
5	Poisson	0.2290	5	Poisson	0.3087
6	M.C of order one	0.0399	6	M.C of order one	0.0787
7	M.C of order two	0.0341	7	M.C of order two	0.0112
8	Eggenberger-Polya	0.0382	8	Eggenberger-Polya	0.0396
Critical value at $\alpha = .05$		0.1024	Critical value at $\alpha = .05$		0.1018

Table 5.7 Results of the Kolmogorov-Smirnov (K-S) Tests for Imphal (2001-2005)

Summer Wet Spells			Summer Dry Spells		
Serial No.	Distributions	K-S Statistic	Serial No.	Distributions	K-S Statistic
1	Discrete Uniform	0.4167	1	Discrete Uniform	0.3333
2	Geometric	0.4416	2	Geometric	0.4663
3	Logarithmic	0.4505	3	Logarithmic	0.4727
4	Neg. Binomial	0.2187	4	Neg. Binomial	0.6678
5	Poisson	0.2415	5	Poisson	0.2466
6	M.C of order one	0.1056	6	M.C of order one	0.0736
7	M.C of order two	0.0435	7	M.C of order two	0.0307
8	Eggenberger-Polya	0.0491	8	Eggenberger-Polya	0.0152
Critical value at $\alpha = .05$		0.1070	Critical value at $\alpha = .05$		0.1064

5.3 Results

This section deals with the comparative results obtained from different statistical models applied to analyze the wet and dry spells frequencies over North East India. In the first phase of this work we have calculated the empirical frequencies of wet and dry spells according to their length. Then the same frequencies have been estimated for each station using the aforesaid theoretical distribution models.

Results of Kolmogorov-Smirnov tests presented in the Table 5.1 to Table 5.7 clearly indicate that apart from the Markov chain of order two (in some cases order 1 also) and Eggenberger-Polya distribution, the rest of the distributions work poorly to represent thc spell frequencies. In case of dry series, Eggenberger-Polya distribution and Markov chain of order two show better results in all seven stations where as Markov chain of order one shows good fit for the stations Mohanbari, Cherrapunji, Guwahati and Imphal. While Eggenberger-Polya distribution gives best fit for the stations North-Bank, Silcoorie, Cherrapunji and Imphal, Markov chain of order two shows best fit for the stations Tocklai, Mohanbari and Guwahati. Summarizing the above experiences, we may conclude that Eggenberger-Polya distribution and Markov chain of order two are competing each other in case of dry spells. In comparison to dry series Markov chain of order two shows better performance in case of wet series. Results of the Kolmogorov-Smirnov tests for Markov chain of order one shows good fit to the observed data in most of the investigated cascs. Like dry spells, Eggenberger-Polya and Markov chain of order two are the best fitting models in case of wet spells also. Markov chain of order two gives best fit to the observed data for four stations and Eggenberger-Polya distribution works better than Markov chain of order two for the rest three stations.

5.4 Conclusion

This section concerns with the critical evaluation of the work carried out. These are listed below

- The best fitting model in a particular station is found to be consistent for estimating both wet and dry spell frequencies. For example, Eggenberger-Polya distribution gives best fit to both wet and dry spells for the station North Bank.

- Eggenberger-Polya distribution and Markov chain of order two (in some cases Markov chain of order one also) models are efficient in fitting the observed data. The other models do not fit at all.

- In case of dry spells (wet spells) Eggenberger-Polya distribution (Markov chain of order two) shows best fit in four out of seven stations.

- Markov chain of order two needs four parameters while Eggenberger-Polya needs only two parameters. Considering the above discussions it can be concluded that Eggenberger-Polya is better than Markov chain of order two and can be more easily used as a theoretical model to estimate the seasonal climatic characterization of precipitation over North-East India.

Chapter 6

Statistical Analysis of Annual Maximum Rainfall based on the Methods of L-moment and LQ-moment

The purpose of this chapter is to determine the best fitting extreme value distribution to describe the annual series of maximum daily rainfall data for the period 1966 to 2007 of nine distantly located stations in North East India. Model parameters are estimated using the method of L-moment and LQ-moment. Finally, goodness of fit test results are compared and generalized logistic distribution is empirically proved to be the most appropriate distribution for describing the annual maximum rainfall series for the majority of the stations in North East India.

6.1 Introduction

Realistic sequences of meteorological variables such as extreme rainfall are key inputs in many hydrologic, ecologic and agricultural models. So there is a pressing need to know the magnitudes of the extreme rainfall events over different parts of the area under study. Moreover, knowledge of spatial and temporal variability of extreme rainfall events is very much useful for the design of dam and hydrological planning. Therefore, study on the statistical modeling of extreme rainfall is very much essential as the statistical model may vary according to the geographical locations of the area considered. So far no rigorous work has been pursued in the North East India to study the annual maximum rainfall events. Considerable efforts have been made in this direction using the annual series of maximum daily rainfall data for the period of 42 years of nine stations in North East India. For this purpose, five three-parameter extreme value distributions viz. Generalized Extreme Value distribution (GEV), Generalized Logistic distribution (GLD), Generalized Pareto distribution (GPD), Lognormal distribution (LN3) and Pearson (P3) distribution are considered. The estimation of the parameters for each distribution has been done using the methods of L-Moment and LQ-Moment independently. The performances of the distributions are evaluated using three goodness of fit tests namely relative root mean square error, relative mean absolute error and probability plot correlation coefficient. Further, L-moment ratio diagram is also used to confirm the goodness of fit for the above five distributions. For the earlier literature concerning the statistical modeling of extreme events, we refer to Section 1.4.

The rest of the chapter is organized as follows. While Section 6.2 introduces a brief specification of data set and the statistical methods used in the present study, Section 6.3 is devoted for a discussion on the results obtained from different statistical models applied to analyze the series of annual maximum rainfall. A concluding remark is given in Section 6.4.

6.2 Data and Methodology

Series of annual maximum daily rainfall data of nine stations in North East India viz Imphal, Mohanbari, Guwahati, Cherrapunji, Pasighat, North Lakhimpur, Silchar, Shillong and Tezpur for a period of 42 years from 1966 to 2007 have been considered for this study. The locations of the nine stations are shown in Figure 1.1. The series of block maxima for annual blocks of daily rainfall data of the aforesaid stations are collected from Regional Meteorological Centre, Guwahati. The set of daily rainfall data is complete for the analysis period and the graphical representation of the data is shown in Figure 6.1.

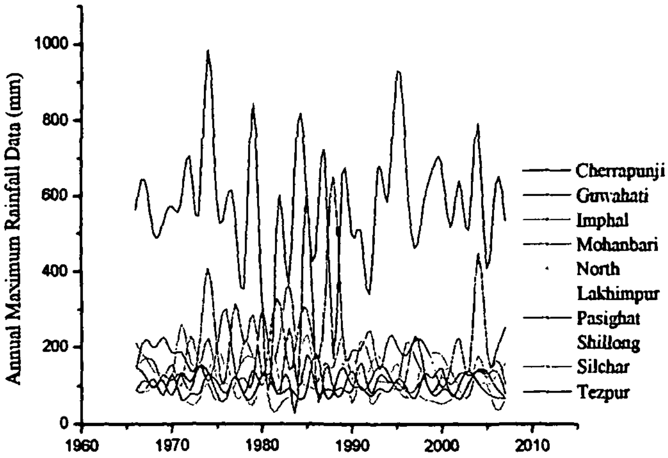

Figure 6.1. Representation of annual maximum rainfall data

In order to describe the behavior of extreme rainfall at a particular area, it is necessary to identify the distribution(s), which best fit the data. In this study, five three-parameter extreme value distributions namely Generalized Extreme Value, Generalized Logistic, Generalized Pareto, Lognormal and Pearson distribution are considered to find the best fitting probability distribution function to extreme rainfall data. The probability density functions of the above distributions along with their quantile functions are exhibited below.

Generalized Extreme Value (GEV) Distribution:

$$f(x) = \frac{1}{\alpha}\left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}-1} \exp\left[-\left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}}\right]$$

where $-\infty < x \leq \xi + \alpha/k$ for $k > 0$ and $\xi + \alpha/k \leq x < \infty$ for $k < 0$.

Quantile function of GEV:

$$Q(F) = \xi + \alpha Q_0(F)$$

where

$$Q_0(F) = [1 - (-\log F)^k]/k. \tag{6.2.1}$$

Generalized Logistic Distribution (GLD):

$$f(x) = \frac{1}{\alpha}\left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}-1}\left[1 + \left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}}\right]^{-2}$$

where $-\infty < x \leq \xi + \alpha/k$ for $k > 0$ and $\xi + \alpha/k \leq x < \infty$ for $k < 0$.

Quantile function of GLD:

$$Q(F) = \xi + \alpha Q_0(F)$$

where

$$Q_0(F) = [1 - \{(1-F)/F\}^k]/k.$$

Generalized Pareto Distribution (GPD):

$$f(x) = \frac{1}{\alpha}\left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}-1}$$

where $\xi < x \leq \xi + \alpha/k$ for $k > 0$ and $\xi \leq x < \infty$ for $k < 0$.

Quantile function of GPD:

$$Q(F) = \xi + \alpha Q_0(F)$$

where

$$Q_0(F) = [1 - (1-F)^k]/k.$$

Lognormal Distribution (LN3):

$$f(x) = \frac{1}{\alpha\sqrt{2\pi}} e^{-\log\left\{1-k\frac{(x-\xi)}{\alpha}\right\} - \frac{1}{2}\left[-\frac{1}{k}\log\left\{1-k\frac{(x-\xi)}{\alpha}\right\}\right]^2}$$

where $-\infty < x \leq \xi + \alpha/k$ for $k > 0$ and $\xi + \alpha/k \leq x < \infty$ for $k < 0$.

Quantile function of LN3:

$$Q(F) = \zeta + \exp(\mu) Q_0(F)$$

where

$$Q_0(F) = \exp[\sigma\Phi^{-1}(F)]$$

and $\Phi^{-1}(.)$ has a standard normal distribution with mean zero and unit variance. Parameters ζ, μ and σ are the standard parameterizations which can be obtained by setting

$$k = -\sigma, \quad \alpha = \sigma e^{\mu}, \quad \xi = \zeta + e^{\mu}.$$

Pearson Distribution (P3):

$$f(x) = \frac{1}{|\beta|\Gamma(\alpha)}\left(\frac{(x-\xi)}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{(x-\xi)}{\beta}\right)\right), \quad -\infty < x < \infty.$$

The quantile function of P3:

$$Q(F) = \mu + \sigma Q_0(F)$$

where

$$Q_0(F) = \frac{2}{\gamma}\left[1 + \frac{\gamma\Phi^{-1}(F)}{6} - \frac{\gamma^2}{36}\right]^3 - \frac{2}{\gamma}$$

and $\Phi^{-1}(.)$ has a standard normal distribution with mean zero and unit variance. Parameters γ, μ and σ are the standard parameterizations which can be obtained by setting

$$\alpha = \frac{4}{\gamma^2}, \quad \beta = \frac{1}{2}\sigma|\gamma|, \quad \xi = \mu - \frac{2\sigma}{\gamma}.$$

To estimate the parameters for each of the aforesaid distributions, methods of L-Moment and LQ-Moment are used independently.

6.2.1 Method of L-Moment

The L-moments (LMOM) were introduced by Sillitto [66] and comprehensively reviewed by Hosking [31] for estimating the parameters of certain statistical distributions. The L-moments are linear functions of the expectations of order statistics and they can be viewed as an alternative system of describing the shapes of probability distributions. The main advantages of using the method of LMOM are that the parameter estimates are more reliable (i.e. smaller mean-squared error of estimation) and are more robust, and are usually computationally more tractable than the conventional moments and maximum likelihood.

Let $X_1, X_2, \ldots, X_n$ be a sample from a continuous distribution function $F(.)$ with quantile function $Q(F)$ and let $X_{1\,n} \leq X_{2\,n} \leq \ldots \leq X_{n\,n}$ denotes the order statistics. Then the rth L-moment λ_r is given by

$$\lambda_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k \begin{pmatrix} r-1 \\ k \end{pmatrix} E(X_{r-k\,r}), \quad r = 1,\ 2, \ldots$$

The details on the estimation of parameters for each of the aforesaid distributions can be found in Hoskıng and Wallis ([32]).

6.2.2 Method of LQ-Moment

Mudholkar and Hutson [50] extended LMOM to a new moment called LQ-moments (LQM) by introducing some quick estimators such as median, trimean or Gastwirth in places of expectations in LMOM. They found that LQM always exists, are often easicr to compute and estimate than LMOM, and in general behave similarly to the LMOM. In fact, in some recent literature such as Shabri et al. [64] it has found that LQM gives better performancc in high quantile estimation as compared to the conventional LMOM.

Analogous to λ_r, the rth LQ-moments ζ_r of X is defined as

$$\zeta_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k \begin{pmatrix} r-1 \\ k \end{pmatrix} \tau_{p,\alpha}(X_{r-k\,r}), \quad r = 1,\ 2, \ldots$$

where $0 \leq \alpha \leq 1/2,\ 0 \leq p \leq 1/2$ and

$$\tau_{p,\alpha}(X_{r-k\,r}) = pQ_{X_{r-k\,r}}(\alpha) + (1-2p)Q_{X_{r-k\,r}}(1/2) + pQ_{X_{r-k\,r}}(1-\alpha)$$

and $Q_X(.)$ is the quantile function. $\tau_{p,\alpha}$ is called the median for $p = 0,\ \alpha = 1$, trimean for $p = 1/4,\ \alpha = 1/4$ and Gastwirth for $p = .3,\ \alpha = 1/3$. In this study trimean based estimator is considered. In parametric estimation the coefficient of skewness and kurtosis play an important role. The LQ skewness (η_3) and LQ kurtosis (η_4) are given by

$$\eta_3 = \zeta_3/\zeta_2 \quad \text{and} \quad \eta_4 = \zeta_4/\zeta_2,$$

respectively.

The LQ moment can be estimated from the sample by estimating the quick estimator

$$\begin{aligned}\hat{\tau}_{p,\alpha}(X_{r-k\,r}) &= p\hat{Q}_{X_{r-k\,r}}(\alpha) + (1-2p)\hat{Q}_{X_{r-k\,r}}(1/2) + p\hat{Q}_{X_{r-k\,r}}(1-\alpha) \\ &= p\hat{Q}_X[B^{-1}_{r-k\,r}(\alpha)] + (1-2p)\hat{Q}_X[B^{-1}_{r-k\,r}(1/2)] \\ &+ p\hat{Q}_X[B^{-1}_{r-k\,r}(1-\alpha)], \;\; 0 \le \alpha \le 1/2, \;\; 0 \le p \le 1/2,\end{aligned}$$

$B^{-1}_{r-k\,r}(\alpha)$ is the αth quantile of a beta random variable with parameters $r-k$ and $k+1$, and $\hat{Q}_X(.)$ denotes the linear interpolation estimator given by

$$\hat{Q}_X(u) = (1-\epsilon)X_{[n'u]\,n} + \epsilon X_{[n'u]+1\,n}$$

where $\epsilon = n'u - [n'u]$ and $n' = n+1$. Then the estimation of the first four sample LQ moments in simplified form are given by

$$\hat{\zeta}_r = \sum_i c_i \hat{Q}_X(u_i).$$

For trimean based functional, the values of c_i and u_i are given in Table 6.1 (cf. [50]).

Table 6.1 Triangular representation for the estimates of the first four LQ moments $\hat{\zeta}_r$ based upon trimean functional are presented in the following table.

$\hat{\zeta}_4$		$\hat{\zeta}_3$		$\hat{\zeta}_2$		$\hat{\zeta}_1$	
c_i	u_i	c_i	u_i	c_i	u_i	c_i	u_i
1/16	0.931	1/12	0.909	1/8	0.866	1/4	0.750
1/8	0.841	1/6	0.794	1/4	0.707	1/2	0.500
-3/16	0.757	-1/6	0.674	-1/4	0.293	1/4	0.250
1/16	0.707	1/12	0.630	-1/8	0.134		
-3/8	0.614	-1/3	0.500				
3/16	0.544	1/12	0.370				
-3/16	0.456	-1/6	0.326				
3/8	0.386	1/6	0.206				
-1/16	0.293	1/12	0.091				
3/16	0.243						
-1/8	0.159						
-1/16	0.069						

In the present study, we evaluate the parameters for each extreme value distribution by solving a nonlinear algebraic equation involving the unknown parameter k. For the completeness of this work, we present the case for Generalized Extreme Value Distribution (GEV). The parameters k, α, ξ for GEV distribution can be estimated using the following relation involving sample LQ skewness

$$\begin{aligned}
\hat{\eta}_3 &= \Big\{\frac{1}{12}Q_0(.909) + \frac{1}{6}Q_0(.794) - \frac{1}{6}Q_0(.674) + \frac{1}{12}Q_0(.630) - \frac{1}{3}Q_0(.5) \\
&\quad + \frac{1}{12}Q_0(.370) - \frac{1}{6}Q_0(.326) + \frac{1}{6}Q_0(.206) + \frac{1}{12}Q_0(.091)\Big\} \Big/ \\
&\quad \Big\{\frac{1}{8}Q_0(.866) + \frac{1}{4}Q_0(.707) - \frac{1}{4}Q_0(.293) - \frac{1}{8}Q_0(.134)\Big\} \\
&= \Big[\frac{1}{12}\{-(.0954)^k\} + \frac{1}{6}\{-(.2307)^k\} - \frac{1}{6}\{-(.3945)^k\} + \frac{1}{12}\{-(.4620)^k\} \\
&\quad - \frac{1}{3}\{-(.6931)^k\} + \frac{1}{12}\{-(.9943)^k\} - \frac{1}{6}\{-(1.1209)^k\} \\
&\quad + \frac{1}{6}\{-(1.5799)^k\} + \frac{1}{12}\{-(2.3969)^k\}\Big] \Big/ \Big[\frac{1}{8}\{-(.1439)^k\} \\
&\quad + \frac{1}{4}\{-(.3467)^k\} - \frac{1}{4}\{-(1.2276)^k\} - \frac{1}{8}\{-(2.0099)^k\}\Big]. \qquad (6.2.2)
\end{aligned}$$

In the last equation we have used the quantile function for GEV distribution given by (6.2.1). In order to solve the above equation for k numerically, we first generate 1000 different values for k in the interval $[-1, 1]$ for suitable step size and those values are used to calculate the right hand side of the equation (6.2.2). If we denote the approximate right hand side by the symbol $\eta_{3:\hat{k}}$ for a particular value of k, then k is chosen in such a way that $|\hat{\eta}_3 - \eta_{3:\hat{k}}|$ is minimum. The estimate of the other two parameters ξ and α of GEV distribution are then given by

$$\begin{aligned}
\hat{\xi} &= \hat{\zeta}_1 - \hat{\alpha}\Big[\frac{1}{4}\hat{Q}_0(1/4) + \frac{1}{2}\hat{Q}_0(1/2) + \frac{1}{4}\hat{Q}_0(3/4)\Big] \\
\hat{\alpha} &= 8\hat{\zeta}_2 \Big/ \Big[2\hat{Q}_0(.707) - 2\hat{Q}_0(.293) + \hat{Q}_0(.866) - \hat{Q}_0(.134)\Big]
\end{aligned}$$

with $\hat{Q}_0(u) = [1 - (-\log u)^{\hat{k}}]/\hat{k}$. Details of the estimated values of the parameters for each distribution using LQM for extreme rainfall are presented in Table 6.4.

6.2.3 Goodness of Fit (GOF)

The tests applied for judging the goodness of fit for the fitted distributions for annual maximum rainfall series are relative root mean squared error (RRMSE), relative mean absolute error(RMAE) and probability plot correlation coefficient(PPCC). While the first two tests involve the assessment on the difference between the observed values and expected values of the assumed distributions, the last one measures the correlation between the ordered values and the corresponding expected values. The formulae for the tests are

$$\text{RRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_{i\,n} - \hat{Q}(F_i)}{x_{i\,n}}\right)^2}$$

$$\text{RMAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_{i\,n} - \hat{Q}(F_i)}{x_{i\,n}}\right|$$

$$\text{PPCC} = \frac{\sum_{i=1}^{n}(x_{i\,n} - \bar{x})\{\hat{Q}(F_i) - \bar{Q}(F)\}}{\sqrt{\sum_{i=1}^{n}(x_{i\,n} - \bar{x})^2}\sqrt{\sum_{i=1}^{n}\{\hat{Q}(F_i) - \bar{Q}(F)\}^2}}.$$

where $x_{i\,n}$ is the observed values of the ith order statistics of a random sample of size n, $\hat{Q}(F_i)$ is the estimated quantile values associated with the ith Gringorten plotting position

$$F_i = \frac{i - .44}{n + .12}, \quad \text{and} \quad \bar{Q}(F) = \frac{1}{n}\sum_{i=1}^{n}\hat{Q}(F_i).$$

The smallest values of RRMSE and RMAE correspond to the best fitting distribution where as in the case of PPCC, the distribution with the computed PPCC closest to 1 indicates the best. In order to confirm the goodness of fit for the above five distributions we additionally applied L-moment ratio diagram. L-moment ratio diagram was first introduced by Hosking ([31]) which can be drawn by plotting L-kurtosis τ_4 as ordinate and L-skewness τ_3 as abscissa. According to Hosking and Wallis ([32]), the simple explicit expressions for τ_4 in terms of τ_3 for the assumed distributions can be written as

$$\tau_4 = \sum_{k=0}^{8} A_k \tau_3^k \qquad (6.2.3)$$

where the coefficients A_k are given in the Table 6.1. Although this is a crude method, it can provide some insights on the selection of the best fitting distribution.

The observed sample L-skewness t_3 for all the nine stations are substituted in place of τ_3 in the expression (6.2.3) to get the estimated L-kurtosis τ_4 for the assumed distributions. These computed values (τ_3, τ_4) for each distributions along with the observed (t_3, t_4) are plotted on the L moment ratio diagram. For a particular station, the distances between (τ_3, τ_4) and (t_3, t_4) for all distributions are compared and evaluated. The distribution corresponding to the smallest distance is considered to be the best.

6.3 Results and Discussion

The extreme rainfall amount can be characterized by mean, standard deviation and coefficient of variation. Table 6.2 provides a quantitative comparison between the rain gauge stations, and it can be concluded that Cherrapunji received the highest mean and standard deviation of annual maximum daily rainfall amount during Indian summer. The coefficients of variations for Mohanbari followed by Pasighat are found to be higher as compared to the stations in other areas. This may indicate that the amounts of extreme rainfall in those two stations are relatively more spread as compared to the other regions of North East India.

The next analysis involves the estimation of parameters for each distribution using LMOM and LQM. The estimated values are given in Table 6.3. The computation is carried out using the software Matlab 6. Subsequent analysis involves selection of the best fitting distribution out of the five candidate distributions. Results for all GOF tests for each station based on L-moment and LQ-moment are presented in Table 6.4. The distribution that is found best at least twice out of the three GOF tests will be selected as the best fitting distribution for both the LMOM and LQM. Then, we summarize the results based on the L-moment ratio diagram (cf. Figure 6.2), LMOM and LQM under the three GOF tests to decide the best fitting distribution for a particular station are given in Table 6.5.

Table 6.2 Polynomial approximations of τ_4 as a function of τ_3

A_i	GPD	GEV	GLD	LN3	P3
A_0	0	0.10701	0.16667	0.12282	0.1224
A_1	0.20196	0.11090	-	-	-
A_2	0.95924	0.84838	0.83333	0.77518	0.30115
A_3	-0.20096	-0.06669	-	-	-
A_4	0.04061	0.00567	-	0.12279	0.95812
A_5	-	-0.04208	-	-	-
A_6	-	0.03763	-	-0.13638	-0.57488
A_7	-	-	-	-	-
A_8	-	-	-	0.11368	0.19383

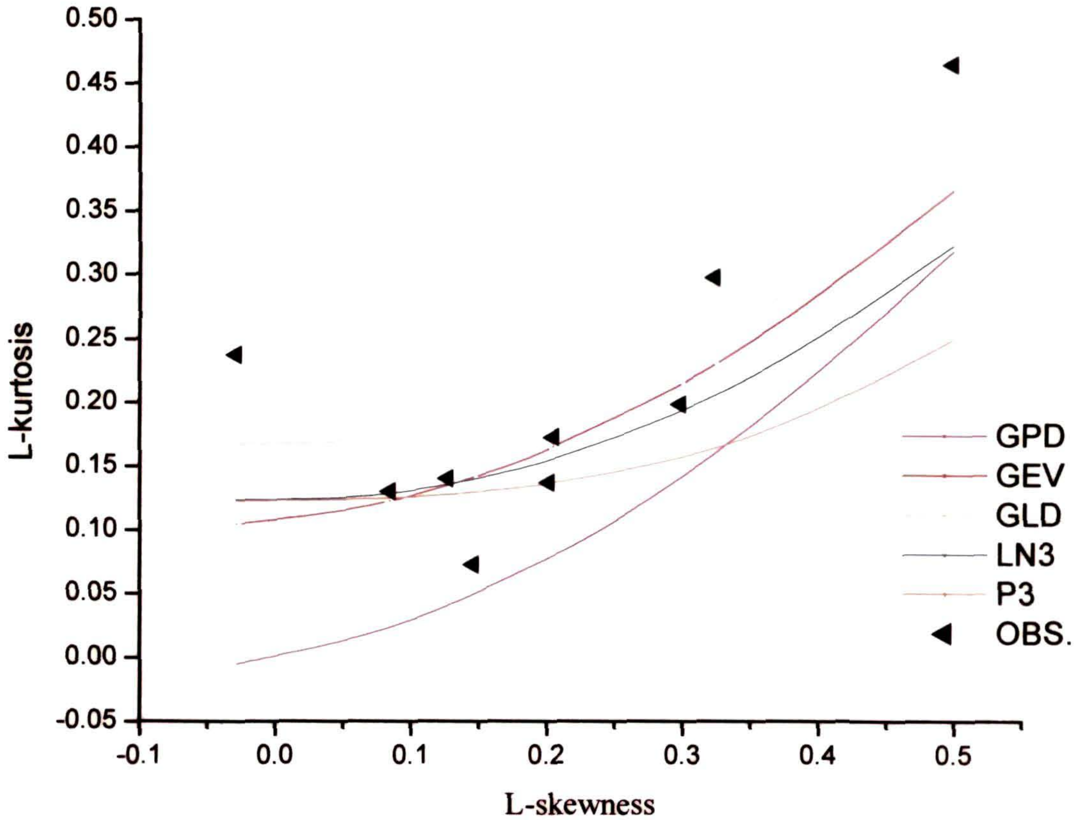

Figure 6.2 L-Moment Ratio Diagram for Annual Maximum Rainfall of 9 stations of North East India

Table 6.3 Main characteristics of the rain gauge stations in North-East India

Stations	Mean	SD	CV
Cherrapunji	573.6167	172.6664	.3010
Guwahati	104.9786	35.2915	.3362
Imphal	82.7619	29.5122	.3566
Mohanbari	142.9476	86.7992	.6072
North Lakhimpur	149.1619	38.1247	.2556
Pasighat	225.3238	98.8148	.4385
Shillong	144.5048	51.3595	.3554
Silchar	153.2524	56.4452	.3683
Tezpur	103.7476	27.3783	.2639

Table 6.4 Estimates of the parameters for each distribution using LMOM and LQM

	GEV		GPD		GLD		LN3		P3	
	LMOM	LQM	LMOM	LQM	LMOM	LQM	LMOM	LQM	LMOM	LQM
Stations	k α ξ	k α ξ	k α ξ	k α ξ	k α ξ	k α ξ	k α ξ	k α ξ	γ σ μ	γ σ μ
Cherrapunji	0 3363	0 3560	1 1173	1 000	0 0285	0 0360	0 0583	0 0640	-0 1749	-0 1920
	172 6430	149 3092	628 8920	419 5162	95 1530	90 1574	168 6404	148 8653	169 0413	149 2093
	518 5045	513 103	276 5963	371 6294	578 0814	582 3670	578 5398	582 4713	573 6167	577 7027
Guwahati	-0 1915	-0 3000	0 0805	0 0360	-0 2985	-0 4160	-0 6242	-0 7280	1 7919	2 0040
	20 8402	21 7337	40 1017	38 11	15 3371	17 3146	26 8691	28 4477	34 8749	37 2045
	88 1314	87 0536	67 8646	69 1262	96 5968	95 5698	95 7198	95 3781	104 9786	106 3616
Imphal	-0 0543	-0 3280	0 3192	-0080	-0 2051	-0 4360	-0 4240	-0 7640	1 2403	2 0880
	22 1224	17 5081	49 4519	30 0019	15 0686	14 1161	26 5802	23 1368	30 0530	31 0654
	68 7411	70 2415	45 2743	55 9885	77 4199	77 1384	76 8654	76 9724	82 7619	86 4243
Mohanbari	-0 4565	-0 3200	-0 3324	0 2920	-0 4995	-0 2960	-1 0912	-0 5160	3 0763	1 4760
	24 7200	23 4385	36 6807	46 7769	20 9966	17 3864	35 4176	28 6435	75 7399	32 9933
	108 5838	114 0436	88 0068	93 2714	119 0238	122 9397	116 5368	122 8082	142 9416	130 4069
North	0 0680	0 0320	0 5486	0 5520	-0 1272	-0 1800	-0 2613	-0 3160	0 7755	0 9280
Lakhimpur	32 4959	32 4931	83 8063	74 0877	20 6730	22 4868	36 5793	37 0933	38 3489	39 1814
	132 4625	130 9820	95 0454	99 9498	144 7546	142 9309	144 3000	142 8181	149 1619	148 7270
Pasighat	-0 2270	-0 0440	0 0217	0 4360	-0 3238	-0 2320	-0 6798	-0 4040	1 9426	1 1760
	53 0678	51 9931	98 4042	111 9100	39 8415	37 1670	69 6121	61 2882	94 6700	66 9654
	179 5084	184 1884	128 0075	136 0804	201 2315	203 6091	198 7062	203 3914	225 3238	215 9816
Shillong	0 1379	0 5840	0 6878	1 0000	-0 0847	0 1840	-0 1736	0 3200	0 5181	-0 9430
	46 7575	56 1169	131 3953	145 2479	28 6249	30 8502	50 6990	50 8995	51 7721	53 8564
	123 1935	128 1228	66 6530	72 1260	140 4849	146 4686	140 0714	146 6097	144 5048	138 3708
Silchar	-0 0489	-0 1840	0 3290	0 2120	-0 2016	-0 3320	-0 4165	-0 5800	1 2192	1 6400
	41 3860	39 5753	93 0847	75 8442	28 1038	30 0186	49 5829	49 4166	55 8246	58 9284
	127 2679	127 9298	83 2108	93 6376	143 4791	143 1093	142 4651	142 8484	153 2524	157 6936
Tezpur	0 0380	0 0367	0 4908	0 5600	-0 1459	-0 1760	-0 3000	-0 3080	0 8877	0 9113
	23 0766	26 0570	57 4158	59 6667	14 9269	17 9989	26 3969	29 6939	28 0887	31 2981
	91 2652	89 7576	65 2339	64 8108	100 0761	99 3349	99 6970	99 2498	103 7476	103 8820

Table 6.5 Outcomes of the GOF tests based on LMOM and LQM methods

Stations	LMOM				LQM			
	RRMSE	RASE	PPCC	BEST	RRMSE	RASE	PPCC	BEST
Cherrapunji	GLD	GLD	GLD	GLD	GLD	GLD	GLD	GLD
Guwahati	GEV	GEV	GLD	GEV	LN3	LN3	GEV	LN3
Imphal	GEV	GEV	P3	GEV	LN3	LN3	GPD	LN3
Mohanbari	GLD	GLD	GLD	GLD	GLD	P3	GLD	GLD
North Lakhimpur	GEV	GEV	GLD	GEV	GEV	P3	GEV	GEV
Pasighat	GLD	GLD	GLD	GLD	GLD	GLD	GLD	GLD
Shillong	P3	GEV	P3	P3	P3	P3	GPD	P3
Silchar	P3	P3	GLD	P3	P3	P3	LN3	P3
Tezpur	GPD	GPD	P3	GPD	P3	P3	P3	P3

Table 6.6 Best fitting distributions based on L-moment ratio diagram, LMOM and LQM methods for all rain gauge stations

Stations	LMOM	LQM	LMOM Ratio Diagram
Cherrapunji	GLD	GLD	GLD
Guwahati	GEV	LN3	LN3
Imphal	GEV	LN3	GEV
Mohanbari	GLD	GLD	GLD
North Lakhimpur	GEV	GEV	LN3
Pachighat	GLD	GLD	GLD
Shillong	P3	P3	LN3
Silchar	P3	P3	P3
Tezpur	GPD	P3	GPD

Table 6.7 Ranking (in descending order) of the distributions for all stations based on methods of LMOM, LQM and L-moment ratio diagram

Ranking	LMOM	LQM	LMOM Ratio Diagram
1	GEV, GLD	GLD, P3	GLD, LN3
2	P3	LN3	GEV, GPD, P3
3	GPD	GEV	-
4	LN3	GPD	-

Under LMOM, it is found that the number of stations identified best using GLD, GEV, LN3, GPD and P3 are 3, 3, 0, 1 and 2 respectively. On the other hand, under LQM, it is found that the number of stations identified best using GLD, GEV, LN3, GPD and P3 are 3, 1, 2, 0 and 3 respectively. Further in the L-moment ratio diagram, number of stations identified best using GLD, GEV, LN3, GPD and P3 are 3, 1, 3, 1 and 1 respectively. This information is summarized in Table 6.7 by ranking them in descending order to show the best fitting distribution for all the stations in North-East India.

6.4 Conclusion

This study reveals that the results of the best fitting distributions may differ for a particular station depending on either LMOM or LQM is used. However, GLD is found to be more consistent in comparison to the other three best fitting distributions. If we consider LMOM, GEV shares the first rank with GLD but fails to perform under LQM and in LMOM ratio diagram. For LQM, P3 is found to be best fitting distribution along with GLD but receives second rank in LMOM and works poorly in case of LMOM ratio diagram. Further, in case of LMOM ratio diagram LN3 distribution holds the first rank with GLD but it is found to be least frequently selected under LMOM methods. From the above discussions, it can be concluded that GLD is the most suitable distribution to describe the annual maximum rainfall in North East India, which also agrees with the result obtained by Zin et al. (2008). But GPD is found to be the least frequently selected distribution. This result differs from the result obtained by Zin et al. (2008) for extreme rainfall in Peninsular Malaysia.

Chapter 7

LH-Moments for Statistical Analysis of Annual Maximum Rainfall

In this chapter, the LH-moments of order zero(L) to order four (L_4) are used to estimate the parameters of three extreme value distributions viz. generalized extreme value distribution, generalized logistic distribution and generalized Pareto distribution are used to estimate the parameters of three extreme value distributions to describe the annual series of maximum daily rainfall data. Finally, it can be revealed that the L level of the generalized Pareto distribution would be appropriate to the majority of the stations for describing the annual maximum rainfall series in North East India.

7.1 Introduction

In the previous chapter, we have discussed five three-parameter extreme value distributions to describe the annual series of maximum daily rainfall data. Model parameters have been estimated using the method of L-moment and LQ-moment independently. Over the years LH-moments have been developed by Wang ([73]) as a generalization of the L-moments with the capacity of a more detailed analysis of annual flood peak data. In his study he concentrated only on the generalized extreme value distribution. Since then LH-moments have been used by several authors in flood frequency analysis. For the details application of LH-moments in flood frequency analysis, we refer to Wang ([73]) and Meshgi et al. ([46], [47]), and the references therein. Although a good number of articles is devoted to the statistical modeling of extreme rainfall using L-moments, there is hardly any literature concerning the use of LH-moments in the statistical modeling of extreme rainfall. Therefore, LH-moments(L to L_4) are used to estimate the parameters of three extreme value distributions viz. generalized extreme value distribution, generalized logistic distribution and generalized Pareto distribution to annual maximum daily rainfall data for the period 1966 to 2007 of nine distantly located stations in North East India. The performances of the distributions are assessed by evaluating the relative bias (RBIAS) and relative root mean square error (RRMSE) of quantile estimates through Monte Carlo simulations. Then the boxplot is used to show the location of the median and the associated dispersion of the data.

The rest of the chapter is organized as follows. While Section 7.2 introduces a brief specification of data set, basics of the LH-moments and statistical tools used in the present study, section 7.3 is devoted to a discussion on the results obtained from different levels of LH-moments for RRMSE and RBIAS values. The chapter ends with a concluding remark.

7.2 Data and Methodology

Series of annual maximum daily rainfall data of nine stations in North East India viz Imphal, Mohanbari, Guwahati, Cherrapunji, Pasighat, North Lakhimpur, Silchar, Shillong and Tezpur for a period of 42 years from 1966 to 2007 have been considered for this study. The locations of the nine stations are shown in Figure 1.1. The series of annual maximum daily rainfall is collected from Regional Meteorological Centre, Guwahati, Assam, India.

7.2.1 Method of LH-Moment

Wang ([73]) introduced the concept of LH-moments as generalization of the L-moment. These are based on linear combination of Higher order statistics Given a sample of size m drawn from a distribution $F(x) = \Pr(X \le x)$, the expectation of the jth smallest variable is given by Hosking ([31]).

$$E[X_{j\,m}] = \frac{m!}{(j-1)!(m-j)!}\int_0^1 x(F)F^{j-1}(1-F)^{m-j}dF.$$

For any probability p, $x(p)$ is the quantile of nonexceedance probability p. Then the LH-moments are defined as

$$\lambda_1^\eta = E[X_{(\eta+1)\,(\eta+1)}]; \qquad \lambda_2^\eta = \frac{1}{2}E[X_{(\eta+2)\,(\eta+2)} - X_{(\eta+1)\,(\eta+2)}]$$

$$\lambda_3^\eta = \frac{1}{3}E[X_{(\eta+3)\,(\eta+3)} - 2X_{(\eta+2)\,(\eta+3)} + X_{(\eta+1)\,(\eta+3)}]$$

$$\lambda_4^\eta = \frac{1}{4}E[X_{(\eta+4)\,(\eta+4)} - 3X_{(\eta+3)\,(\eta+4)} + 3X_{(\eta+2)\,(\eta+4)} - X_{(\eta+1)\,(\eta+4)}].$$

When the order of LH-moments, $\eta = 0$, LH-moments are equivalent to L-moments As η increases, LH-moments reflect more and more the characteristics of the upper part of the distributions and larger events in data. LH-moments are called L_1 moments, L_2 moments, ... for $\eta = 1, 2, \ldots$, respectively LH-moments can be normalized to define LH coefficient of variation, skewness, and kurtosis, respectively, as $\tau_2^\eta = \frac{\lambda_2^\eta}{\lambda_1^\eta}$, $\tau_3^\eta = \frac{\lambda_3^\eta}{\lambda_2^\eta}$, $\tau_4^\eta = \frac{\lambda_4^\eta}{\lambda_2^\eta}$.

For a given ranked sample, $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$, the sample estimates of the LH-moments can be estimated by (cf. Wang [73])

$$\hat{\lambda}_1^\eta = \frac{1}{{}^nC_{\eta+1}}\sum_{i=1}^{n} {}^{i-1}C_\eta x_{(i)}, \quad \hat{\lambda}_2^\eta = \frac{1}{2}\frac{1}{{}^nC_{\eta+2}}\sum_{i=1}^{n}({}^{i-1}C_{\eta+1} - {}^{i-1}C_\eta\ {}^{n-i}C_1)x_{(i)}$$

$$\hat{\lambda}_3^\eta = \frac{1}{3}\frac{1}{{}^nC_{\eta+3}}\sum_{i=1}^{n}({}^{i-1}C_{\eta+2} - 2\ {}^{i-1}C_{\eta+1}\ {}^{n-i}C_1 + {}^{i-1}C_\eta\ {}^{n-i}C_2)x_{(i)}$$

$$\hat{\lambda}_4^\eta = \frac{1}{4}\frac{1}{{}^nC_{\eta+4}}\sum_{i=1}^{n}({}^{i-1}C_{\eta+3} - 3\ {}^{i-1}C_{\eta+2}\ {}^{n-i}C_1 + 3\,{}^{i-1}C_{\eta+1}\ {}^{n-i}C_2 - {}^{i-1}C_\eta\ {}^{n-i}C_3)x_{(i)},$$

where

$${}^mC_j = \binom{m}{j} = \frac{m!}{j!(m-j)!}$$

is the number of combination of any j item from m items and is zero when $j > m$.

L-moments are a linear transformation of Probability weighted moments (PWM) (cf. [31]) and LH-moments are a generalization of L-moments (cf. [73]). Therefore LH-moments are a linear combination of Higher order PWMs, and the relationships between the LH-moments and Normalized PWMs are given by Wang ([73]) as

$$\lambda_1^\eta = B_\eta \tag{7.2.1}$$

$$\lambda_2^\eta = \frac{1}{2!}(\eta+2)[B_{\eta+1} - B_\eta] \tag{7.2.2}$$

$$\lambda_3^\eta = \frac{1}{3!}(\eta+3)[(\eta+4)B_{\eta+2} - 2(\eta+3)B_{\eta+1} + (\eta+2)B_\eta] \tag{7.2.3}$$

$$\begin{aligned}\lambda_4^\eta = \frac{1}{4!}(\eta+4)[&(\eta+6)(\eta+5)B_{\eta+3} - 3(\eta+5)(\eta+4)B_{\eta+2} + 3(\eta+4)(\eta+3)B_{\eta+1} \\ &-(\eta+3)(\eta+2)B_\eta],\end{aligned} \tag{7.2.4}$$

where,

$$B_r = \int_0^1 x(F)F^r dF \Big/ \int_0^1 F^r dF = (r+1)\int_0^1 x(F)F^r dF = (r+1)\beta_r \tag{7.2.5}$$

and β_r is the normalized PWM and is the standard PWM (cf. Greenwood et al. [30]).

In order to describe the behavior of extreme rainfall at a particular area, it is necessary to identify the distribution(s), which best fit the data. In this study, three extreme value distributions namely Generalized Extreme Value, Generalized Logistic and Generalized Pareto distributions are considered to find the best fitting probability distribution function to extreme rainfall data. To estimate the parameters for each of the aforesaid distributions, methods of LH-Moments are used. Although the pdf and Quantile functions for the aforesaid distributions are given in the previous chapter, we again recall them for the simplicity of the exposition.

LH-moments for Generalized Extreme Value (GEV) Distribution:

The Probability density function of GEV is given by

$$f(x) = \frac{1}{\alpha}\left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}-1} \exp\left[-\left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}}\right]$$

where $-\infty < x \leq \xi + \alpha/k$ for $k > 0$ and $\xi + \alpha/k \leq x < \infty$ for $k < 0$.

Quantile function of GEV:

$$Q(F) = \xi + \alpha Q_0(F), \quad Q_0(F) = [1 - (-\log F)^k]/k.$$

The PWMs of GEV developed by Hosking ([31]) is

$$B_r = \left\{ \xi + \frac{\alpha}{k}[1 - \Gamma(1+k)(r+1)^{-k}] \right\}. \tag{7.2.6}$$

Then combining the identities (7.2.1)-(7.2.4) with equation (7.2.6) leads to a system of equations involving the parameters α, ξ and k. In the evaluation of the parameters, the sample LH-moments ($\hat{\lambda}_1^\eta$, $\hat{\lambda}_2^\eta$, $\hat{\lambda}_3^\eta$, $\hat{\lambda}_4^\eta$) may be used directly. So we propose first estimating the shape parameter k by numerically solving the following non-linear equation

$$\hat{\tau}_3^\eta = \frac{\lambda_3^\eta}{\lambda_2^\eta} = \frac{(\eta+3)}{3(\eta+2)} \left[\frac{-(\eta+4)(\eta+3)^{-k} + 2(\eta+3)(\eta+2)^{-k} - (\eta+2)(\eta+1)^{-k}}{-(\eta+2)^{-k} + (\eta+1)^{-k}} \right] \tag{7.2.7}$$

in the interval $[-1, 1]$. In order to solve equation (7.2.7) for k numerically, we first generate 1000 different values for k in the interval $[-1, 1]$ for suitable step size and those values are used to calculate the right hand side of the equation (7.2.7). If we denote the approximate right hand side by the symbol $\tau_{3\hat{k}}^\eta$ for a particular value of k, then k is chosen in such a way that $|\hat{\tau}_3^\eta - \tau_{3\hat{k}}^\eta|$ is minimum. The estimate of the other two parameters α and ξ are then given by

$$\alpha = \frac{2!k\hat{\lambda}_2^\eta}{(\eta+2)\Gamma(k+1)} \frac{1}{[-(\eta+2)^{-k} + (\eta+1)^{-k}]}$$
$$\xi = \hat{\lambda}_1^\eta - \frac{\alpha}{k}[1 - (\eta+1)^{-k}\Gamma(k+1)].$$

LH-moments for Generalized Logistic Distribution (GLD):

The Probability density function of GLD is given by

$$f(x) = \frac{1}{\alpha}\left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}-1}\left[1 + \left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}}\right]^{-2}$$

where $-\infty < x \le \xi + \alpha/k$ for $k > 0$ and $\xi + \alpha/k \le x < \infty$ for $k < 0$.

Quantile function of GLD:

$$Q(F) = \xi + \alpha Q_0(F), \; Q_0(F) = [1 - \{(1-F)/F\}^k]/k.$$

Then substituting above information in (7.2.5), we have

$$B_r = \left\{\xi + \frac{\alpha}{k}\left[1 - (r+1)\beta(r-k+1, k+1)\right]\right\},$$

$\beta(.,.)$ is the standard beta function. The shape parameter k for GLD distribution can be computed using the following relation

$$\hat{\tau}_3^\eta = \frac{\lambda_3^\eta}{\lambda_2^\eta} = \frac{(\eta+3)}{3(\eta+2)}\frac{L}{M}, \tag{7.2.8}$$

where

$$\begin{aligned} L &= -(\eta+4)(\eta+3)\beta(\eta-k+3, k+1) \\ &\quad +2(\eta+3)(\eta+2)\beta(r-k+2, k+1) - (\eta+2)(\eta+1)\beta(\eta-k+1, k+1) \\ M &= -(\eta+2)\beta(\eta-k+2, k+1) + (\eta+1)\beta(\eta-k+1, k+1). \end{aligned}$$

Arguing as in estimating k for GEV distribution, the approximate value of k can be found numerically solving the equation (7.2.8) in the interval $[-1, 1]$ for GLD distribution. The estimate of the other two parameters α and ξ are then given by

$$\alpha = \frac{2!k\hat{\lambda}_2^\eta}{(\eta+2)}\Big(-(\eta+2)\beta(\eta-k+2, k+1) + (\eta+1)\beta(\eta-k+1, k+1)\Big)^{-1}$$

$$\xi = \hat{\lambda}_1^\eta - \frac{\alpha}{k}\big(1 - (\eta+1)\beta(\eta-k+1, k+1)\big).$$

LH-moments for Generalized Pareto Distribution (GPD):

The Probability density function of GPD is given by

$$f(x) = \frac{1}{\alpha}\left\{1 - k\frac{(x-\xi)}{\alpha}\right\}^{\frac{1}{k}-1}$$

where $\xi < x \le \xi + \alpha/k$ for $k > 0$ and $\xi \le x < \infty$ for $k < 0$.

Quantile function of GPD:

$$Q(F) = \xi + \alpha Q_0(F),\ Q_0(F) = [1 - (1-F)^k]/k.$$

Substitute above information in (7.2.5) to have the following PWMs of GPD distribution

$$B_r = \xi + \frac{\alpha}{k}\Big[1 - (r+1)\beta(r+1, k+1)\Big].$$

The shape parameter k for GPD distribution can be computed using the following relation

$$\hat{\tau}_3^{\eta} = \frac{\lambda_3^{\eta}}{\lambda_2^{\eta}} = \frac{(\eta+3)}{3(\eta+2)}\frac{P}{Q}, \tag{7.2.9}$$

where

$$\begin{aligned} P &= -(\eta+4)(\eta+3)\beta(\eta+3,k+1) \\ &\quad +2(\eta+3)(\eta+2)\beta(\eta+2,k+1) - (\eta+2)(\eta+1)\beta(\eta+1,k+1) \\ Q &= -(\eta+2)\beta(\eta+2,k+1) + (\eta+1)\beta(\eta+1,k+1). \end{aligned}$$

We then calculate the approximate value of k satisfying (7.2.9) arguing as in the case for GEV distribution. The estimate of the other two parameters α and ξ are then given by

$$\begin{aligned} \alpha &= \frac{2!k\hat{\lambda}_2^{\eta}}{(\eta+2)}\Big(-(\eta+2)\beta(\eta+2,k+1) + (\eta+1)\beta(\eta+1,k+1)\Big)^{-1} \\ \xi &= \hat{\lambda}_1^{\eta} - \frac{\alpha}{k}\big(1-(\eta+1)\beta(\eta+1,k+1)\big). \end{aligned}$$

Parameters for the GEV, GLD and GPD distributions are estimated for each nine stations using the methodology as stated in this section. For the simplicity of the exposition, only the results for the station Cherapunji are presented in Table 7.1. The details of the estimated parameters are presented in Appendix.

7.2.2 Monte Carlo Simulations

The next step in our analysis is to evaluate the performance of different LH-moments level of the GEV, GLD and GPD distribution. For this purpose the Monte Carlo simulations have been carried out to evaluate the LH-moments levels of GEV, GPD and GLD distributions in terms of their capabilities when estimating quantiles of specific recurrence intervals. In this case the parameters are estimated from the observed data and then the values of the parameters are used to generate random samples of same size. In each simulation a total of 10,000 samples are generated for a particular recurrence

Table 7.1: Regional parameters of region Cherapunji for the GEV, GLD and GPD distributions, for different levels of the LH-moments

Region	Distribution	η	ξ	α	k
Cherapunji	GEV	0	518.4767	172.6151	.3360
		1	521.5445	130.5284	.0960
		2	530.5818	109.5680	-.0040
		3	534.2021	104.1172	-.0280
		4	534.1806	104.1266	-.0280
	GLD	0	578.0034	95.1574	.0280
		1	572.9681	86.8092	-.0600
		2	573.9340	80.5648	-.1040
		3	574.2426	79.9061	-.1080
		4	573.1956	81.4028	-.1000
	GPD	0	287.7761	571.6812	1.000
		1	405.7723	253.9468	.3720
		2	452.1337	172.9941	.1560
		3	470.2451	148.5069	.0840
		4	477.2774	140.3607	.0600

interval(RI). The simulations have been conducted for each of the nine rain gauge stations separately and for RI=2, 5, 10, 20, 50, 100 years. Then the parameters for each random sample are estimated using the technique describe in Subsection 3.2. Finally, these information are used to estimate the quantile functions ($Q_{s(m)}$) for each simulated sample.

The criteria for the selection of a particular PDF at a particular LH-moment level would be based on the minimum error produced when simulated and calculated quantiles are compared for a number of recurrence intervals. Two of the more commonly error function used in such cases are the relative root mean square error (RRMSE) and relative bias (RBIAS) represented by

$$\text{RRMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\frac{Q_{s(m)} - Q_c}{Q_c}\right)^2} \tag{7.2.10}$$

$$\text{RBIAS} = \sum_{m=1}^{M}\left(\frac{Q_{s(m)} - Q_c}{Q_c}\right). \tag{7.2.11}$$

For a particular recurrence interval, M is the total numbers of samples, $Q_{s(m)}$ represents the simulated quantiles of the mth sample and Q_c is the calculated quantiles from the observed data. The minimum RRMSE and RBIAS values and their associated variability are used to select the most suitable PDF at a particular LH-moment level. For this purpose boxplots are used.

The Monte Carlo simulations conducted in this study involves the following steps.

1. Select a set of GEV parameters (in case of the selected rain gauge stations, at site estimated parameters are used).

2. Estimate the quantiles for the RI of interest by using the selected parameters .

3. Using a random number generator, generate a data series, using the selected GEV parameters in step one (in case of the selected rain gauge stations, generated sample size is same as the size of the observed maxima series).

4. Fit the GEV distribution to the generated samples by using LH-moment of order zero.

5. Estimate quantiles for the same RI of step 2.

6. Repeat step 3 to 5 for 10,000 times.

7. Estimate the RRMSE and RBIAS of the quantiles, by taking the quantile estimates in step 2 as the true value.

8. Repeat the procedure from 1 to 7 for L_1, L_2, L_3, L_4 moments.

9. Repeat the procedure from 1 to 8 for GLD and GPD distributions.

Table 7.2 RRMSE values for different recurrence intervals of GEV, GPD and GLD distributions for regions Cherrapunji and Guwahati

Region	Distribution	η	2	5	10	20	50	100
Cherapunji	GEV	0	.0513	**.0407**	**.0405**	**.0437**	**.0551**	**.0676**
		1	.0429	.0491	.0518	.0619	.0837	.1075
		2	.0390	.0472	.0565	.0701	.0993	.1288
		3	**.0368**	.0505	.0580	.0702	.1032	.1327
		4	.0398	.0499	.0599	.0745	.1022	.1350
	GPD	0	.0554	**.0419**	**.0337**	**.0343**	**.0403**	**.0448**
		1	.0452	.0530	.0500	.0543	.0642	.0779
		2	.0373	.0531	.0613	.0662	.0872	.1076
		3	**.0352**	.0509	.0612	.0715	.0953	.1216
		4	.0359	.0536	.0623	.0715	.0987	.1225
	GLD	0	.0450	**.0423**	**.0471**	**.0573**	**.0773**	**.0925**
		1	.0421	.0448	.0520	.0706	.0969	.1261
		2	.0404	.0462	.05775	.0714	.1079	.1426
		3	**.0389**	.0473	.0555	.0711	.1085	.1468
		4	.0419	.0493	.0564	.0740	.1054	.1446
Guwahati	GEV	0	**.0457**	**.0634**	.0827	.1128	.1753	.2282
		1	.0479	.0669	**.0816**	.1150	.1653	.2243
		2	.0505	.0704	.0894	**.1094**	**.1648**	**.2181**
		3	.0502	.0720	.0850	.1114	**.1648**	.2240
		4	.0497	.0714	.0924	.1155	.1661	.2320
	GPD	0	.0517	**.0704**	**.0809**	**.0974**	**.1285**	**.1597**
		1	.0477	.0747	.0851	.1039	.1519	.1838
		2	**.0449**	.0763	.0908	.1081	.1525	.2028
		3	.0459	.0740	.0957	.1170	.1627	.2280
		4	.0459	.0738	.0907	.1111	.1618	.2118
	GLD	0	**.0460**	**.0615**	**.0105**	.1234	.1942	.2579
		1	.0488	.06288	.0842	.1114	.1797	.2407
		2	.0517	.0658	.0815	**.1094**	**.1669**	**.2267**
		3'	.0502	.0703	.0861	.1125	.1752	.2301
		4	.0536	.0677	.0895	.1142	.1729	.2575

Table 7.3 RBIAS values for different recurrence intervals of GEV, GPD and GLD distributions for regions Cherrapunji and Guwahati

Region	Distribution	η	2	5	10	20	50	100
Cherapunji	GEV	0	.0007	-.00086	-.0016	**.0015**	**.0027**	.0042
		1	.0014	-.0026	-.0035	-.0033	.0017	.0062
		2	.0020	**-.00035**	-.0018	-.0020	-.0037	**.0016**
		3	.0006	-.00065	**.0003**	-.0045	-.0096	-.0056
		4	**.0002**	.0023	.0020	-.0056	-.0098	-.0054
	GPD	0	-.0090	-.0016	.0072	.0122	.0214	.0246
		1	.0016	-.0021	-.0018	-.0028	.0035	.0062
		2	.0006	.0009	-.0041	-.0050	**-.0023**	.0016
		3	**.0002**	.0007	**-.0012**	**-.0026**	-.0049	**.0007**
		4	-.0014	**.0003**	-.0024	-.0048	-.0103	-.0061
	GLD	0	.0005	-.0007	**.00008**	**-.0028**	.0046	.0071
		1	.0018	-.0024	-.0024	-.0031	.0030	.0052
		2	.0017	**.0004**	-.0044	-.0059	-.0041	-.0020
		3	.0020	.0015	-.0007	-.0085	**-.0016**	**-.0011**
		4	**-.00002**	.0005	-.0021	-.0063	-.0031	-.0021
Guwahati	GEV	0	.0024	**-.0012**	**-.0033**	-.0118	**.0022**	.0055
		1	.0032	-.0028	-.0054	-.0090	-.0023	**.0011**
		2	.0022	.0017	-.0070	-.0118	-.0113	-.0095
		3	**.00079**	.0038	-.0040	**-.0088**	-.0117	-.0110
		4	-.0018	.0041	-.0045	-.0114	-.0170	-.0280
	GPD	0	.0017	**-.0010**	-.0040	-.0056	-.0057	**.0049**
		1	**.0015**	.0022	.0036	**-.0030**	-.0031	-.0057
		2	.0022	.0024	.0028	-.0118	**.0004**	.0006
		3	-.0034	.0046	**.0003**	-.0072	-.0255	-.0150
		4	.0016	.0038	.0034	-.0109	.0130	-.0150
	GLD	0	.0016	-.0033	-.0088	-.0111	**-.0037**	-.0058
		1	.0028	-.0011	-.0052	-.0100	-.0079	-.0036
		2	.0023	**-.0003**	-.0054	-.0120	-.0101	**-.0035**
		3	**.0004**	.0038	-.0076	**-.0012**	-.0151	-.0204
		4	-.0020	.0018	**.00001**	-.0089	-.0141	-.0155

7.3 Results and Discussions

The present section is devoted to determine the best fitting extreme value distribution to describe the annual series of maximum daily rainfall data for the period 1966 to 2007 of nine distantly located stations in North East India. The performances of the distributions are assessed by evaluating the relative bias (RBIAS) and relative root mean square error (RRMSE) of quantile estimates through Monte Carlo simulations. Selection of the most suitable distribution function is based on the smallest calculated values of RRMSE and RBIAS at different levels of η.

7.3.1 RRMSE and RBIAS values by different PDF's

The efficiency of GEV, GLD and GPD distributions under RRMSE and RBIAS test for different recurrence intervals (i.e. $2, 5, 10, 20, 50, 100$) at different levels of η has been discussed in this section. Table 7.2 describes the RRMSE values for regions Cherrapunji and Guwahati. As the results of Table 7.2 for Cherrapunji, the minimum RRMSE values of GEV, GPD and GLD distributions appears at $\eta = 3$ for recurrence interval 2, and is at $\eta = 0$ for the remaining recurrence intervals. As for region Guwahati, it is observed that the minimum RRMSE value of the GEV distribution is at $\eta = 0$ for recurrence intervals 2 and 5, $\eta = 1$ for 10, and $\eta = 2$ for rest of the recurrence intervals; the minimum RRMSE values of GPD distribution appears at $\eta = 2$ for 2, and is at $\eta = 0$ for the remaining recurrence intervals; the minimum RRMSE values of GLD distribution appears at $\eta = 0$ for 2, 5 and 10, and is at $\eta = 2$ for the remaining intervals. Therefore a significant conclusion can not be drawn from the RRMSE values and hence we have omitted the details about the RRMSE values for rest of the stations.

Let us turn our discussion to RBIAS values for regions Cherrapunji and Guwahati. Table 7.3 describes the RBIAS values for different recurrence intervals of GEV, GLD and GPD distributions at different levels of η. As in the case of RRMSE values, RBIAS values require further analysis, so that logical selection of the LH-moments levels can be made for individuals PDF's for each station. For this purpose the box plots can be used as a tool for grouping of results based on statistical properties, as we will be discussed next.

7.3.2 Boxplots for better illustration of the RRMSE and RBIAS results

Box plot is a widely used graphical tool introduced by Tukey ([72]). It is a simple plot of five quantities namely the minimum value, the lower quantile ($q_{0.25}$), the median ($q_{0.5}$), the upper quantile ($q_{0.75}$) and maximum value. This provides the location of the median and associated dispersion of the data at specific probability levels. Then, in those cases where it is difficult to reach a decisive conclusion among several levels of variability of computed values (as in the case of'RRMSE and RBIAS tests) box plots can provide useful information. Figures 7.1-7.9, provide the associated box plots of relative positions of RRMSE values of LH-moments of GEV, GPD and GLD distributions for the nine stations considered in this study.

The criterion for selecting a suitable LH-moments level is based on the minimum achieved median RRMSE or RBIAS values, as well as the minimum dispersion in the median RRMSE or RBIAS values, indicated by both ends of the box plot. It is noted that a smaller median dispersion in RRMSE or RBIAS values would indicate better integration of the LH-moments levels, so it should also be used as selection criterion. As illustrated by Figure 7.1, for region Cherrapunji, the GPD distribution at $\eta = 0$ level produces the minimum median and dispersion in RRMSE. While figures 7.2, 7.3, 7.4, 7.5, 7.6, 7.8 illustrates similar results, Figures 7.7 and 7.9 indicate that the L_1 level of the GPD distribution has minimum dispersion in RRMSE. Figure 7.10 illustrating the relative positions of'RBIAS values of LH moments of GEV, GPD and GLD distributions for station Cherranpunji. As illustrated by Figure 7.10, almost all LH moment levels of the corresponding PDF's have produced very low RBIAS values. As a result it is rather difficult to select one particular distribution function. Similar conclusion can be drawn for the stations North Lakhimpur, Shillong and Silchar as illustrated by the Figures 7.14, 7.16 and 7.17, respectively. The L_2 level of GPD and L level of GEV distribution produces the minimum median RBIAS value of $-.0005$ for station Guwahati, but the RBIAS dispersion for L_2 level of GPD distribution (indicated by both ends of the box plot) is quite high compare to the L level of GEV distribution as illustrated by Figure 7.11. Figure 7.12 shows RBIAS values for station Imphal and clearly L_1 level of GPD distribution produces the minimum median RBIAS value of $-.00015$ and minimum dispersion.

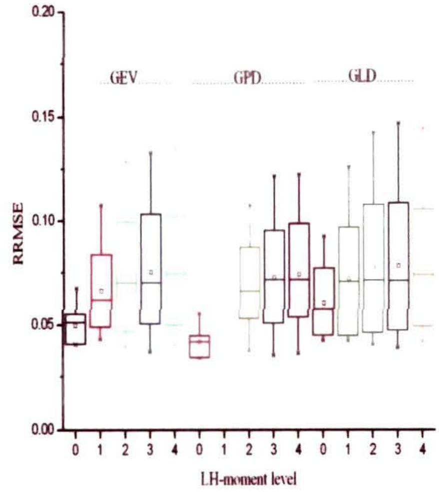

Figure 7.1 Box plots of RRMSE values, Cherrapunji

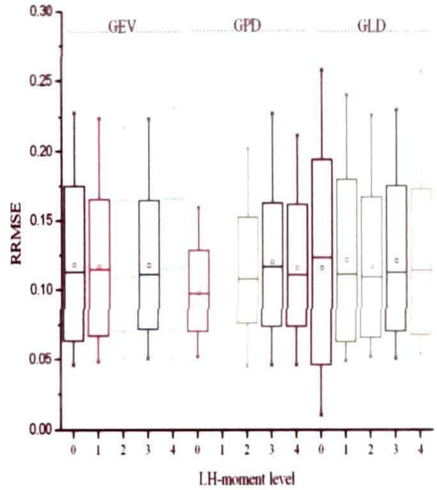

Figure 7.2 Box plots of RRMSE values, Guwahati

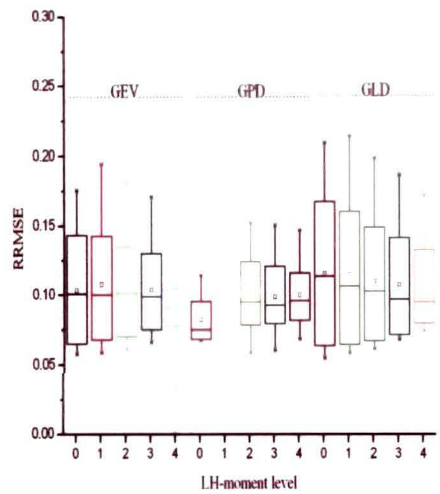

Figure 7.3 Box plots of RRMSE values, Imphal

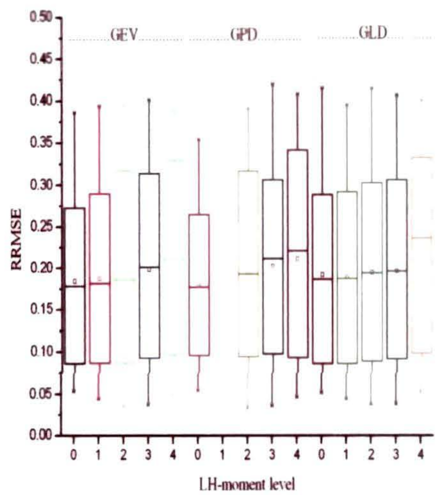

Figure 7.4 Box plots of RRMSE values, Mohanbari

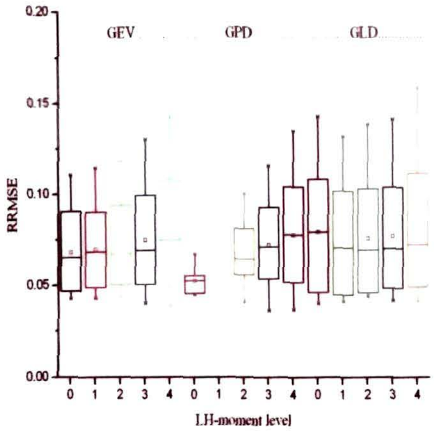

Figure 7.5 Box plots of RRMSE values, North Lakhimpur

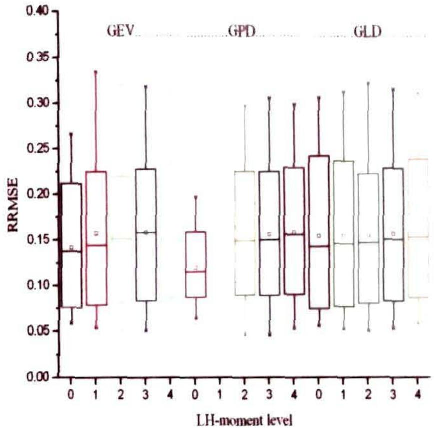

Figure 7.6 Box plots of RRMSE values, Pasighat

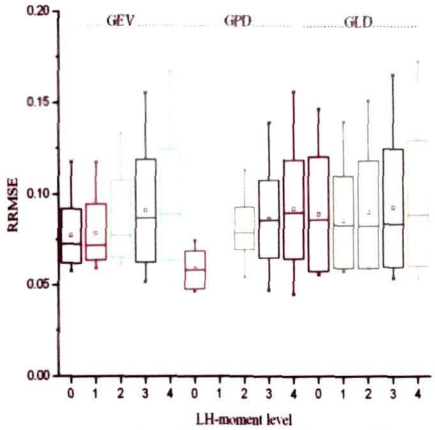

Figure 7.7 Box plots of RRMSE values, Shillong

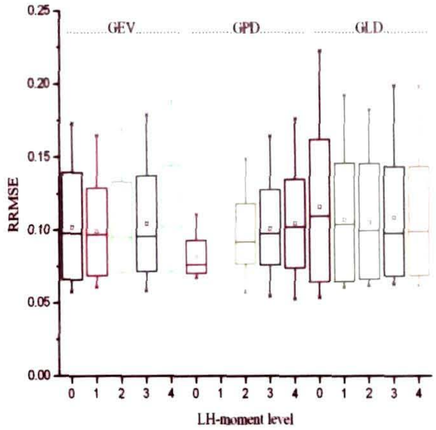

Figure 7.8 Box plots of RRMSE values, Silchar

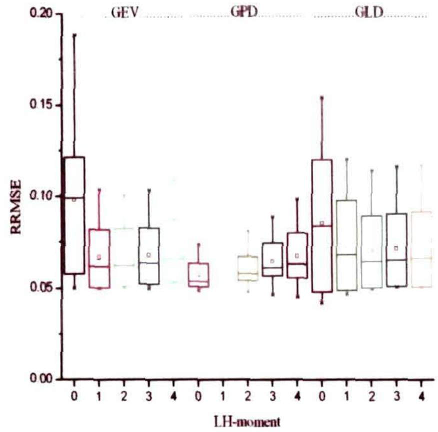

Figure 7.9 Box plots of RRMSE values, Tezpur

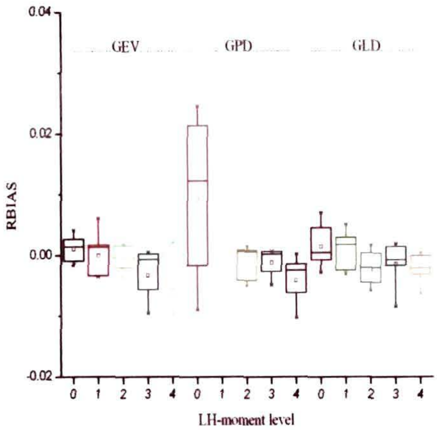

Figure 7.10. Box plots of RBIAS values, Cherrapunji

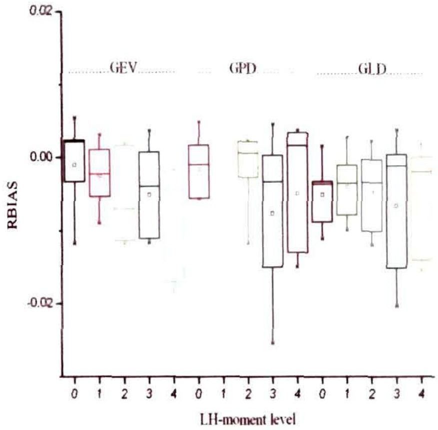

Figure 7.11. Box plots of RBIAS values, Guwahati

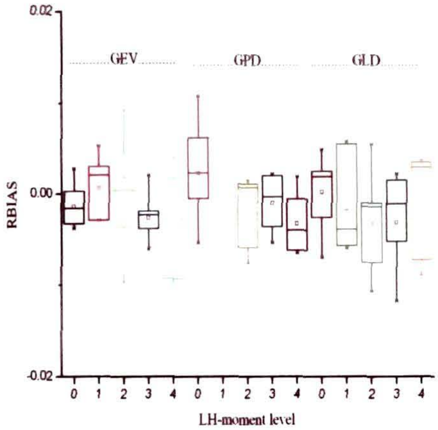

Figure 7.12. Box plots of RBIAS values, Imphal

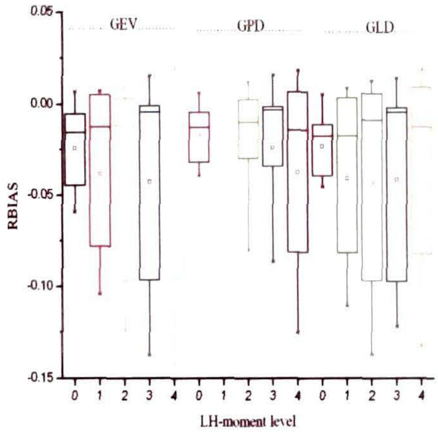

Figure 7.13. Box plots of RBIAS values, Mohanbari

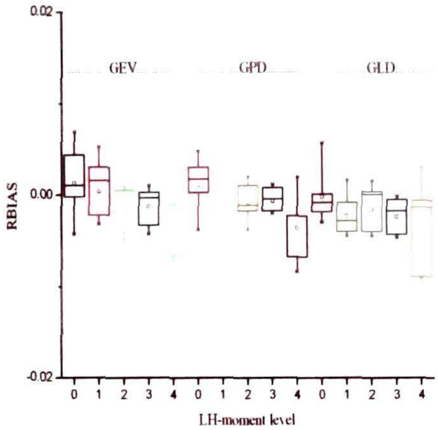

Figure 7.14. Box plots of RBIAS values, North-Lakhimpur

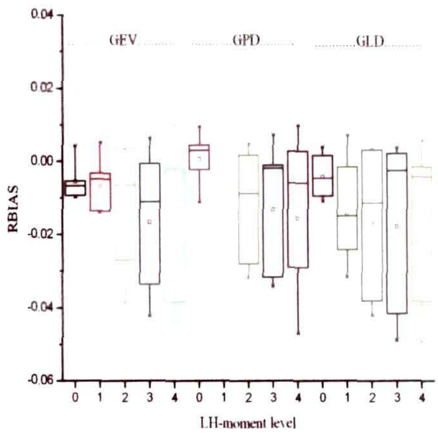

Figure 7.15. Box plots of RBIAS values, Pasighat

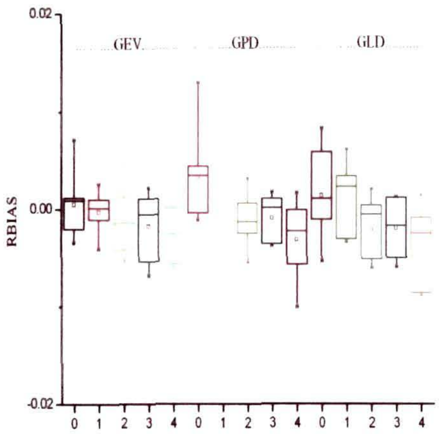

Figure 7.16. Box plots of RBIAS values. Shillong

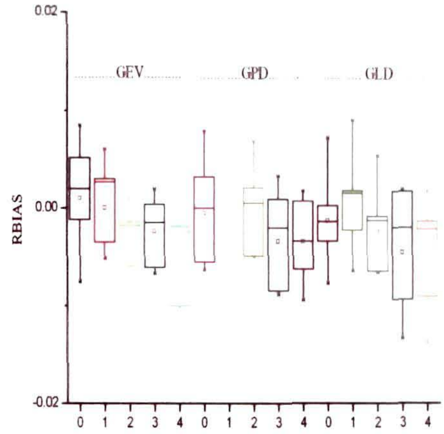

Figure 7.17 Box plots of RBIAS values. Silchar

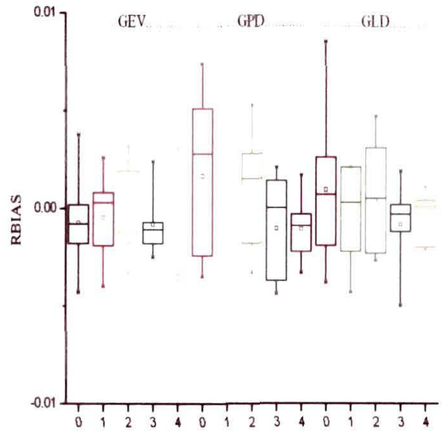

Figure 7.18 Box plots of RBIAS values. Tezpur

Figure 7.13 shows RBIAS values for station Mohanbari and clearly L_2 level of GPD distribution produces the minimum median RBIAS value of $-.01115$ and minimum dispersion. Similar observation can be made for station Tezpur as illustrated by Figure 7.18. Figure 7.15 shows RBIAS values for station Pasighat and clearly L level of GPD distribution produces the minimum median RBIAS value of $-.001545$ and minimum dispersion.

7.4 Conclusion

This study is intended to model maximum/extreme rainfall in the North East India. Any crop producing potentiality of an area depends primarily on the prevailing climate and soil conditions. A fore-knowledge of rainfall pattern is of immense help not only to farmers, but also to the authorities concerned with planning of irrigation schemes. With this in mind this study is being carried to examine what kind of distribution would be appropriate for extreme rainfall. If the best fitting distribution is known for a particular station, one would be able to predict the return value of this extreme rainfall event at a specific time in the future.

It is important to note from the earlier studies (cf. Zalina et al. (2002), Zin et al. (2008) and Kysely et al. (2007)) on the statistical modeling of extreme rainfall that the best fitting probability distribution may vary according to the geographical locations of the area considered and the method used to estimate the parameters. Although theoretical result (cf. Coles (2007)) suggest that for block maxima the appropriate class is generalized extreme value distribution. This study reveals that generalized Pareto distribution would be appropriate for describing the annual maximum rainfall series in North East India when the distributions are fitted using LH-moments. More precisely, zero level of LH-moments of GPD is found to be superior to the majority of the stations in comparison to the other higher levels of LH-moments. Further, higher levels of the LH-moments can also be used to obtain improve estimate values of extreme rainfall for some stations in North East India.

Bibliography

[1] Aronica, G., Cannarozzo, M. and Noto, L., Investigating the changes in extreme rainfall series recorded in an urbanised area, *Water Science & Technology* **45**, 49-54 (2002).

[2] Akaike, H., *Information Theory and an Extension of the Maximum Likelihood Principle*, Second International Symposium on Information Theory, Ed. B. N. Petrov & F. Csóki, (Budapest: Akademiai Kiado, 267-281, 1972).

[3] Akaike, H., *Canonical Correlation Analysis of time Series and the use of an Information Criterion*, Systems Identification: Advances and Case Studies, Ed. R. K. Mehra & D. G. Lainiotie, (New York Academic Press, USA, 27-96, 1976).

[4] Akaike, H., A new look at the statistical identification model, *IEEE Trans. Auto. Control* **AC-19**, 716-723 (1974).

[5] Aksoy, H., Use of Gamma distribution in hydrological analysis, *Turk J. Engi. Environ. Sci.* **24**, 419-428 (2000).

[6] Barger, G. L. and Thom, H. C. Ş., Evaluation of drought hazard, *Agronomy Journal* **41**, 519-526 (1949).

[7] Baloutsos, G. and Koutsoyiannis, D., Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences, *Nat. Haz.* **22**, 31-51 (2000).

[8] Berger, A. and Goossens, Chr., Persistence of wet and dry spells at Uccle (Belgium), *J. Climatol.* **3**, 21-34 (1983).

[9] Bhargava, P. N., Narain, P., Aneja, K. G. and Asha, P., A study of the occurrence of rainfall in Raipur District with the help of Markov chain model, *Journal of Indian Society of Agricultural statistics* **24**, 197-204 (1972).

[10] Bruhn, J. A , Fry, W. E. and Fick, G. W , Simulation of daily weather data using theoretical probability distributions, *J. Appl. Met.* **19**, 1029-1036 (1980).

[11] Coles, S., *An Introduction to Statistical Modeling of Extreme Values*, (Springer Series in Statistics, London, 2007).

[12] Cox, D. R. and Millar, H. D., *The Theory of Stochastic Process*, (Chapman & Hall, London, 1977).

[13] Das, P. J., *Rainfall Regime of Northeast India: A Hydrometeorological Study with Special Emphasis on the Brahmaputra Basin*, (Unpublished Ph.D. Thesis. Gauhati University, 2004).

[14] Deka, S., Kakaty, S. C. and Borah, M., Statistical modeling of rainfall data: Markov chain approach, *Journal of Empirical Research in Social Science* **3**, 77-82 (2008).

[15] Deka, S., Kakaty, S. C. and Borah, M., Fitting of probability distributions for daily rainfall data of North-East India, *Contribution to Applied and Mathematical Sciences* **4**, 42-52 (2009).

[16] Deka, S., Kakaty, S. C. and Borah, M., Use of probability distributions for the analysis of daily rainfall data of North East India, *Mausam* **59**, 518-527 (2008).

[17] Deka, S., Kakaty, S. C and Borah, M., Statistical modeling of wet and dry spell frequencies over North-East India, *Journal of Applied and Natural Science* **2**, 42-47 (2010).

[18] Deka, S., Kakaty, S. C. and Borah, M., Distributions of annual maximum rainfall series of North-East India, *European Water* **27/28**, 3-14 (2010).

[19] Deka, S., Kakaty, S. C. and Borah, M., Statistical analysis of annual maximum rainfall in North-East India: An application of LH-moments, under revision to *Theoretical and Applied Climatology.*

[20] Dubrovsky, M., Creating daily weather series with use of the weather generator, *Environmetrics* **8**, 409-424 (1997).

[21] Fisher, R. A., The influence of rainfall on the yield of wheat at Rothamsted, *Phil. Trans. Roy. Stat. Soc. London. B* **213**, 89-142 (1924).

[22] Gabriel, K. R. and Neumann, J., A Markov chain model for daily occurence at Tel Aviv, *Quart. J. R. met. Soc.* **88**, 90-95 (1962).

[23] Gabriel, K. R. and Neumann J., On a distribution of weather cycles by length, *Quart. J. R. Met. Soc* **83**, 375-380 (1972).

[24] Gates, P. and Tong, H., On Markov chain modeling to some weather data, *J. App. Meteorol.* **15**, 1145-1151 (1976).

[25] Geng, S., A simple method for generating daily rainfall data, *Agricultural and Forest Meteorology* **36**, 363-376 (1986).

[26] Giuseppe, E. D., Vento, D., Epifani C. and Esposito S., Analysis of dry and wet spells from 1870 to 2000 in four Italian sites, *Geophysical Research Abstracts* **7**, 1-6 (2005).

[27] Good I. J., The serial test for sampling numbers and other tests for randomness, *Proc. Camb. Phil. Soc.* **49**, 276-284 (1953).

[28] Gore P. G. and Thapliyal V., Occurrence of dry and wet weeks over Maharastra, *Mausam* **51**, 24-38 (2000).

[29] Goswami, D. C., and Das, P. J., The Brahmaputra river, India: The eco- hydrological context of water use in one of world's most unique river systems, *Ecologist*

Asia, Special issue on large dams in northeast India- Rivers, forests, people and power **11**, 9-14 (2003).

[30] Greenwood, J. A., Landwehr, J. M., Matalas N. C. and Wallis, J. R., Probability weighted moments: definition and relation to parameters of distribution expressible in inverse form, *Water Resour. Res.* **15**, 1049-1054 (1979).

[31] Hosking, J. R. M., L-moments: analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc., Ser. B* **52**, 105-124 (1990).

[32] Hosking, J. R. M. and Wallis, J. R., *Regional Frequency Analysis: An Approach Based on L-Moments*, (University press, Cambridge, 1997).

[33] Hewa, G. A., *et al.*, Generalized extreme value distribution fitted by LH moments for low flow frequency analysis, *Water Resourse Res.* **43**, W 06301, doi:10.1029, 2006WR004913 (2007).

[34] Recorded proceedings of the two day 'Climate Change and Vulnerability of Mountain Ecosystems in the Eastern Himalayan Region, North-East India & Bhutan Stakeholders Workshop', 11-12 March, 2008, Shillong, Organised by International Centre for Integrated Mountain Development Kathmandu, Nepal, 2008.

[35] *Summary for Policymakers in Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. D. Solomon, M. Qin, Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller (eds.), (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007).

[36] Katz, R. W., Computing probabilities associated with the Markov chain model for precipitation, *J. Appl. Meteorol.* **13**, 953-954 (1974).

[37] Katz, R. W., On some criteria for estimating the order of a Markov chain, *Technometrics* **23**, 243-249 (1981).

[38] Keeping, E. S., *Introduction to Statistical Inference*, (D. Van Nostrand Co., Inc., Princeton, N.J., 1962).

[39] Kitagawa, G., On the use of AIC for the detection of outliers, *Technometrics* **21**, 193-199 (1979).

[40] Kotz, S. and Nadarajah, S., *Extreme Value Distributions: Theory and Applications*, (Imperial College Press, London, 2000).

[41] Koutsoyiannis, D. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records, *Hydrological Sciences Journal* **49**, 591-610 (2004).

[42] Kysely, J. and Picek, J., Probability estimates of heavy precipitation events in a flood-prone central-European region with enhanced influence of Mediterranean cyclones, *Adv. Geosci.* **12**, 43-50 (2007).

[43] Matyasovszky, I. and Dobi, I., Methods for analysis of time series of precipitation data using Markov chains, *IdoÈjaÂraÂs* **93**, 276-288 (1989).

[44] Massey, J. and Frank, Jr., The Kolmogorov test for goodness of fit, *JASA* **46**, 68-78 (1951).

[45] Medhi, J., A Markov chain for the occurrence of wet and dry days, *Ind. J. Met. Hydro. & Geophys.* **27**, 431-435 (1976).

[46] Meshgi, A. and Khalili, D., Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. I. A re-visit to regional homogeneity, *Stoch. Environ. Res. Risk Assess* **23**, 119-135 (2009).

[47] Meshgi, A. and Khalili, D., Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. II. Development of LH-moments parameters for the generalized Pareto and generalized logistic distributions, *Stoch. Environ. Res. Risk Assess* **23**, 137-152 (2009).

[48] Mırza, M M Q , Warrıck, R A , Erıcksen, N J and Kenny, G J , Trends and persıstence ın precıpıtatıon ın the Ganges, Brahmaputra and Meghna rıver basıns, *Hydrologıcal Scıences- Journal-des Hydrologıques* **43**, 845-858 (1998)

[49] Mooley, D A , Gamma dıstrıbutıon probabılıty model for Asıan summer monsoon monthly raınfall, *Monthly Weather Revıew* **101**, 160-176 (1973)

[50] Mudholkar, G S and Hutson, A D , LQ-moments analogs of L-moments, *J Stat Plan Inference* **71**, 191-208 (1998)

[51] Muralıdharan, K and Lathıka, P , Statıstıcal modelıng of raınfall data usıng modıfied Weıbull dıstrıbutıon, *Mausam* **56**, 765-770 (2005)

[52] Nadarajah, S and Wıthers, C S , Evıdence of trend ın return levels for daıly raınfall ın New Zealand, *Journal of Hydrology*, **39**, 155-166 (2001)

[53] Nadarajah, S , Extremes of daıly raınfall ın west central Florıda, *Clımatıc Change*, **69**, 325-342 (2005)

[54] Nobılıs, F , Dry spells ın the Alpıne country Austrıa, *J Hydrol* **88**, 235-251 (1986)

[55] Otsu, K , Horıgome, M and Kıtagawa, G , On the predıctıon and stochastıc control of shıp's motıon, *Proc 2nd IFAC/IFIP Sympos* , Ed M Pıtkın, J J Roche & T J Wıllıams, Washıngton, DC, USA, 69-76 (1976)

[56] Otomo, T Nakagawa, T and Akaıke, H , Statıstıcal approach to computer control of cement rotary kıln, *Automatıca* **8**, 35-48 (1972)

[57] Pal, S K , *Statıstıcs for Geoscıentısts Technıques and Applıcatıons*, (Concept Publıshıng Company, New Delhı, 1998)

[58] Racsko, P , Szeıdl, L and Semenov, L , A serıal approach to local stochastıc weather models, *Ecologıcal Modelıng* **57**, 27-41 (1991)

[59] Rakhecha, P. R. and Soman, M. K., Trends in the annual extreme rainfall events of 1 to 3 days duration over India, *Theor. Appl. Climatol.*, **48**, 227-237 (1994).

[60] Rao, A. R. and Hamed, H. K., *Flood Frequency Analysis*, (CRC, Boca Raton, FL. USA, 2000).

[61] Richardson, C. W., Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resources Research*, **17**, 182-190 (1981).

[62] Sakamoto, Y. and Akaike, H., Analysis of cross classified data by AIC, *Ann. Inst. Statist. Math. B* **30** , 185-197 (1978).

[63] Salas, J. P., Delluer, J. W., Yevjevich, V. and Lane, W. L.,*Applied Modelling of Hydrologie Time Series*, (Water Resources Publication, Fort Collins, Colorado, USA, 1980).

[64] Shabri, A. and Jemain, A. A., LQ-moment: application to generalized extreme value, *Journal of Applied Sciences* **7**, 115-120 (2007).

[65] Sharma, T. C., A Markov-Weibull rain-sum model for designing rain water catchment systems, *Water Resources Management* **10**, 147-162 (1996).

[66] Sillitto, G. P., Interrelations between certain linear systematic statistics of samples from any continuous population, *Biometrika* **38**, 377-382 (1951).

[67] Simpson, J., Use of the Gamma distribution in single cloud rainfall analysis, *Monthly Weather Review* **100**, 309-312 (1972).

[68] Stendinger, J. R., Vogel, R. M. and Georgiou, E. F., *Frequency Analysis of Extreme Events*, (Handbook of Hydrology, McGraw Hill, New York, 1992).

[69] Stern, R. D. and Coe, R., A model fitting analysis of daily rainfall data, *Journal of the Royal Statistical Society. Series A (General)* **147**, 1-34 (1984).

[70] Tolika, K. and Maheras, P., Spatial and temporal characteristics of wet spells in Greece, *Theor. Appl. Climatol.* **81**, 71-85 (2005).

[71] Tong, H., Determination of the order of a Markov chain by Akaike's Information Criterion, *Journal of Applied Probability* **12**, 488-497 (1975).

[72] Tukey, J. W., *Expolratory data analysis*, (Addision-Wesley, Reading, 1977).

[73] Wang, Q. J., LH-moments for statistical analysis of extreme events, *Water Resourse Res.* **33**, 2841-2848 (1997).

[74] Wantuch, W., Mika, J. and Szeidl, L., Modelling of wet and dry spells with mixture distributions, *Meteorology and Atmospheric Physics*, **73**, 245-256 (2000).

[75] Wilks, D., Adapting stochastic weather generation algorithm for climate change studies, *Climate Change* **22**, 67-84 (1992).

[76] Zalina, M. D., Desa, M. N., Nguyen, V. T. V. and Hashim, M. K., Selecting a probability distribution for extreme rainfall series in Malaysia, *Waters Sci. Technol*, **45**, 63-68 (2002).

[77] Zhao, L. C., Dorea, C. C. and Goncalves, C. R., On determination of the order of a Markov chain, *Statist. Inferen. Stoch. Proc.* **4**, 273-282 (2001).

[78] Zin, W. Z. W. and Jemain, A. Á., The best fitting distribution of annual maximum rainfall in Peninsular Malaysia based on method of L-moment and LQ-moment, *Theor. Appl. Climatol.* **96**, 337-344 (2008).

Appendix

Appendix A1:

Illustrates the values of Likelihood statistic ${}_k\eta_M$ used for determining $R(k)$ in Chapter 3.

Appendix A2:

Here we have presented the estimates of the parameters for the distributions used in Chapter 5.

Appendix A3:

Performance of each distribution based on LMOM and LQM under different GOF tests described in Chapter 6 are presented here.

Appendix A4:

Details of the parameters for each distribution used in Chapter 7 are presented. Further, RRMSE and RBIAS values for different recurrence intervals of GEV, GPD and GLD distribution are illustrated.

Table A1. 1 Year-wise Likelihood statistic for North Bank

Year	$_0\eta_1$	$_0\eta_2$	$_0\eta_3$	$_1\eta_2$	$_1\eta_3$	$_2\eta_3$	$_3\eta_3$
1986	16.7276	16.8175	19.8498	0.0899	3.1222	3.0323	0
1987	17.7543	22.8454	28.7556	5.0911	11.0013	5.9102	0
1988	41.1159	42.1816	50.9963	1.0657	9.8804	8.8147	0
1989	19.1132	24.3546	32.6416	5.2414	13.5284	8.287	0
1990	21.1622	21.2464	27.8429	0.0842	6.6807	6.5965	0
1991	6.2446	12.6413	21.1247	6.3967	14.8801	8.4834	0
1992	4.6952	9.8015	11.2007	5.1063	6.5055	1.3992	0
1993	18.9827	23.6082	25.9841	4.6255	7.0014	2.3759	0
1994	20.5584	23.2673	26.4903	2.7089	5.9319	3.223	0
1995	28.3186	36.5763	44.1681	8.2577	15.8495	7.5918	0
1996	23.5193	24.5876	27.8354	1.0683	4.3161	3.2478	0
1997	17.4515	19.047	26.7143	1.5955	9.2628	7.6673	0
1998	42.0208	45.0825	50.9531	3.0617	8.9323	5.8706	0
1999	12.1565	16.5651	20.6124	4.4086	8.4559	4.0473	0
2000	4.8409	8.9134	15.9362	4.0725	11.0953	7.0228	0
2001	6.8661	7.3759	9.4704	0.5098	2.6043	2.0945	0
2002	22.3583	22.542	26.9619	0.1837	4.6036	4.4199	0
2003	17.034	19.485	23.7815	2.451	6.7475	4.2965	0
2004	25.857	29.9857	30.2572	4.1287	4.4002	0.2715	0
2005	12.9343	16.5079	20.4342	3.5736	7.4999	3.9263	0
df	1	3	7	2	6	4	0

Table A1. 2 Year-wise Likelihood statistic for Silcoorie

Year	$_0\eta_1$	$_0\eta_2$	$_0\eta_3$	$_1\eta_2$	$_1\eta_3$	$_2\eta_3$	$_3\eta_3$
1986	13.6074	17.4178	20.1958	3.8104	6.5884	2.778	0
1987	19.5835	24.193	26.6532	4.6095	7.0697	2.4602	0
1988	12.4091	14.1805	17.3324	1.7714	4.9233	3.1519	0
1989	22.3025	25.9143	28.4774	3.6118	6.1749	2.5631	0
1990	16.4772	16.568	19.8544	0.0908	3.3772	3.2864	0
1991	12.2873	13.8512	14.4625	1.5639	2.1752	0.6113	0
1992	14.7334	27.0228	27.3544	12.2894	12.621	0.3316	0
1993	15.4073	23.3662	27.737	7.9589	12.3297	4.3708	0
1994	20.7559	22.7451	28.1871	1.9892	7.4312	5.442	0
1995	9.5292	10.1109	12.0423	0.5817	2.5131	1.9314	0
1996	31.6156	38.7471	41.6852	7.1315	10.0696	2.9381	0
1997	13.3732	16.4926	19.6338	3.1194	6.2606	3.1412	0
1999	29.5452	32.9294	38.1983	3.3842	8.6531	5.2689	0
2001	10.4032	17.8828	21.5898	7.4796	11.1866	3.707	0
2002	18.682	24.8816	25.6622	6.1996	6.9802	0.7806	0
2003	19.1264	26.033	30.485	6.9066	11.3586	4.452	0
2004	6.322	8.4268	17.6647	2.1048	11.3427	9.2379	0
2005	40.5932	41.056	58.7494	0.4628	18.1562	17.6934	0

Table A1. 3 Year-wise Likelihood statistic for Mohanbari

Year	$_0\eta_1$	$_0\eta_2$	$_0\eta_3$	$_1\eta_2$	$_1\eta_3$	$_2\eta_3$	$_3\eta_3$
1993	11.5804	15.4773	20.3891	3.8969	8.8087	4.9118	0
1994	13.7926	16.7273	19.4926	2.9347	5.7	2.7653	0
1995	11.3016	13.3507	21.322	2.0491	10.0204	7.9713	0
1996	22.9515	26.0365	27.724	3.085	4.7725	1.6875	0
1997	14.7569	15.6268	21.6405	0.8699	6.8836	6.0137	0
1999	5.0693	16.1724	19.5773	11.1031	14.508	3.4049	0
2001	9.4339	11.3226	11.3835	1.8887	1.9496	0.0609	0
2002	20.2853	20.6465	21.6799	0.3612	1.3946	1.0334	0
2003	10.2603	19.981	22.2287	9.7207	11.9684	2.2477	0
2004	17.1833	19.6397	29.5802	2.4564	12.3969	9.9405	0
2005	7.8452	9.1387	21.4747	1.2935	13.6295	12.336	0
2006	19.5462	23.8717	31.0646	4.3255	11.5184	7.1929	0

Table A1. 4 Year-wise Likelihood statistic for Cherrapunji

Year	$_0\eta_1$	$_0\eta_2$	$_0\eta_3$	$_1\eta_2$	$_1\eta_3$	$_2\eta_3$	$_3\eta_3$
2001	32.8704	36.0696	39.296	3.1992	6.4256	3.2264	0
2002	16.7328	21.1327	30.5769	4.3999	13.8441	9.4442	0
2003	18.8185	24.0015	30.7548	5.183	11.9363	6.7533	0
2004	7.2269	7.6927	14.7354	0.4658	7.5085	7.0427	0
2005	25.6524	30.5269	35.1522	4.8745	9.4998	4.6253	0

Table A1. 5 Year-wise Likelihood statistic for Guwahati

Year	$_0\eta_1$	$_0\eta_2$	$_0\eta_3$	$_1\eta_2$	$_1\eta_3$	$_2\eta_3$	$_3\eta_3$
2001	6.0487	9.2597	12.9944	3.211	6.9457	3.7347	0
2002	16.8919	18.2388	24.4165	1.3469	7.5246	6.1777	0
2003	8.4173	10.4512	31.846	2.0339	23.4287	21.3948	0
2004	6.2384	6.48483	12.63423	0.24643	6.39583	6.1494	0
2005	13.5985	21.8917	32.7865	8.2932	19.188	10.8948	0

Table A1. 6 Year-wise Likelihood statistic for Imphal

Year	$_0\eta_1$	$_0\eta_2$	$_0\eta_3$	$_1\eta_2$	$_1\eta_3$	$_2\eta_3$	$_3\eta_3$
2001	12.6588	12.8072	18.2847	0.1484	5.6259	5.4775	0
2002	26.2281	43.9825	46.9413	17.7544	20.7132	2.9588	0
2003	16.3229	22.8556	30.9987	6.5327	14.6758	8.1431	0
2004	10.1968	13.461	23.5471	3.2642	13.3503	10.0861	0
2005	19.2993	20.7245	26.7122	1.4252	7.4129	5.9877	0

Table A1. 7 Year-wise Likelihood statistic for Tocklai

Year	${}_0\eta_1$	${}_0\eta_2$	${}_0\eta_3$	${}_1\eta_2$	${}_1\eta_3$	${}_2\eta_3$	${}_3\eta_3$
1986	4.7431	7.5144	11.3875	2.7713	6.6444	3.8731	0
1987	10.4907	13.627	15.5623	3.1363	5.0716	1.9353	0
1988	15.7402	17.0761	20.4774	1.3359	4.7372	3.4013	0
1989	13.7658	14.4378	14.8393	0.672	1.0735	0.4015	0
1990	6.4112	12.7139	24.4431	6.3027	18.0319	11.7292	0
1991	2.8172	3.499	11.678	0.6818	8.8608	8.179	0
1992	2.8611	4.8666	9.3829	2.0055	6.5218	4.5163	0
1993	4.0348	9.7398	13.6476	5.705	9.6128	3.9078	0
1994	6.6991	11.8554	16.2017	5.1563	9.5026	4.3463	0
1995	9.5724	13.0751	15.1907	3.5027	5.6183	2.1156	0
1996	9.1044	18.8687	22.1645	9.7643	13.0601	3.2958	0
1997	9.0384	10.4853	15.08526	1.4469	6.04686	4.59996	0
1998	3.3479	9.9012	18.8008	6.5533	15.4529	8.8996	0
1999	5.4121	15.0774	19.4214	9.6653	14.0093	4.344	0
2000	4.8565	7.0168	16.1448	2.1603	11.2883	9.128	0
2001	1.967	6.5032	8.3613	4.5362	6.3943	1.8581	0
2002	10.4907	11.7623	17.7916	1.2716	7.3009	6.0293	0
2003	9.2501	10.0349	13.6387	0.7848	4.3886	3.6038	0
2004	3.2669	7.9666	8.5525	4.6997	5.2856	0.5859	0
2005	4.5919	12.0652	14.2872	7.4733	9.6953	2.222	0

Table A2.1 Estimates of the parameters for North Bank

Distribution	Parameters	
	Summer Dry Spells	Summer Wet Spells
Uniform	a=-1 b=6	a=-2 b=8
Geometric	p=0.28349	p=0.22901
Logarithmic	θ=0.80526	θ=0.87502
Neg. Binomial	n=2 p=0.50638	n=1 p=0.33914
Poisson	λ=2.5274	λ=3.3666
Eggenberger-Polya	m=1.5274 d=2.2677	m=2.3667, d=3.1946

Table A2. 2 Estimates of the parameters for Tocklai

Distribution	Parameters	
	Summer Dry Spells	Summer Wet Spells
Uniform	a=0 b=5	a=-1 b=7
Geometric	p=0.32067	p=0.25431
Logarithmic	θ=0.74113	θ=0.84566
Neg. Binomial	n=5 p=0.72213	n=3 p=0.51236,
Poisson	λ=2.1185	λ=2.9322
Eggenberger-Polya	m=1.1185 d=1.6229	m=1.9322, d=1.9618

Table A2. 3 Estimates of the parameters for Silcoorie

Distribution	Parameters	
	Summer Dry Spells	Summer Wet Spells
Uniform	a=-1 b=5	a=-3 b=11
Geometric	p=0.31121	p=0.19371
Logarithmic	θ=0.75901	θ=0.90884
Neg. Binomial	n=3 p=0.59723	n=1 p=0.24163
Poisson	*λ=2.2133*	*λ=4.1625*
Eggenberger-Polya	m=1.2133 d=2.0544	m=3.1625, d=4.4473

Table A2. 4 Estimates of the parameters for Mohanbari:

Distribution	Parameters	
	Summer Dry Spells	Summer Wet Spells
Uniform	a=-3 b=10	a=-2 b=9
Geometric	p=0.22037	p=0.22222
Logarithmic	θ=0.88401	θ=0.88212
Neg. Binomial	n=1 p=0.24784	n=1 p=0.2543
Poisson	λ=3.5379	λ=3.5
Eggenberger-Polya	m=2.5379, d=4.6247	m=2.5000, d=4.5052

Table A2. 5 Estimates of the parameters for Cherrapunji.

Distribution	Parameters	
	Summer Dry Spells	Summer Wet Spells
Uniform	a=0 b=4	a=-6 b=20
Geometric	p=0.33993	p=0.12623
Logarithmic	θ=0.70096	θ=0.95571
Neg. Binomial	n=24 p=0.92743	n=1 p=.120779
Poisson	λ=1.9417	λ=6.9223
Eggenberger-Polya	m=.9417, d=1.2232	m=5.9223, d=8.677

Table A2. 6 Estimates of the parameters for Guwahati.

Distribution	Parameters	
	Summer Dry Spells	Summer Wet Spells
Uniform	a=0 b=5	a=-1 b=6
Geometric	p=0.29421	p=0.2623
Logarithmic	θ=0.78837	θ=0.83535
Neg. Binomial	n=4 p=0.6463	n=2 p=0.50571
Poisson	λ=2.3989	λ=2.8125
Eggenberger-Polya	m=1.3987, d=1.6533	m=1.8125, d=2.0684

Table A2. 7 Estimates of the parameters for Imphal.

Distribution	Parameters	
	Summer Dry Spells	Summer Wet Spells
Uniform	a=-1 b=7	a=-3 b=8
Geometric	p=0.26942	p=0.25275
Logarithmic	θ=0.8257	θ=0.84761
Neg. Binomial	n=1 p=0.4236	n=1 p=.748511
Poisson	λ=2.7117	λ=2.9565
Eggenberger-Polya	m=1.7117 d=2.7400	m=1.9565 d=5.0086

Table A3. 1 Values of all GOF tests for each station based on LMOM and LQM methods

Stations	GEV		GPD		GLD		LN3		P3	
	LMOM	LQM	LMOM	LQM	LMOM	LQM	LMOM	LQM	LMOM	LQM
	RRMSE RASE PPCC	RRMSE RASE PPCC	RRMSE RASE PPCC	RRMSE RASE PPCC	RRMSE RASE PPCC	RRMSE RASE PPCC	RRMSE RASE PPCC	RRMSE RASE PPCC	RRMSE RASE PPCC	RRMSE RASE PPCC
Cherrapunji	0 1113 0 0586 0 9822	0 0509 0 0060 0 9819	0 2111 0 1067 0 9428	0 0969 0 0102 0 9446	0 0806 0 0370 0 9917	0 0287 0 0036 0 9917	0 1001 0 0522 0 9859	0 0456 0 0054 0 9859	0 1003 0 0523 0 9859	0 0458 0 0054 0 9859
Guwahati	0 0364 0 0300 0 9896	0 0162 0 0034 0 9907	0 0466 0 0377 0 9781	0 0137 0 0031 0 9810	0 0378 0 0311 0 9914	0 0195 0 0038 0 9889	0 0373 0 0306 0 9874	0 0135 0 0029 0 9892	0 0417 0 0341 0 9825	0 0135 0 0031 0 9840
Imphal	0 0382 0 0281 0 9877	0 0336 0 0053 0 9458	0 0619 0 0471 0 9865	0 0349 0 0054 0 9751	0 0439 0 0296 0 9829	0 0373 0 0056 0 9297	0 0385 0 0287 0 9883	0 0318 0 0052 0 9633	0 0406 0 0319 0 9890	0 0352 0 0054 0 9735
Mohanbari	0 0862 0 0659 0 9722	0 0383 0 0064 0 8863	0 1000 0 0784 0 9628	0 0443 0 0077 0 8101	0 0847 0 0641 0 9732	0 0360 0 0060 0 9155	0 0972 0 0769 0 9651	0 0394 0 0066 0 8741	0 1259 0 1013 0 9375	0 0204 0 0017 0 8596
North Lakhimpur	0 0281 0 0242 0 9930	0 0087 0 0021 0 9932	0 0534 0 0376 0 9782	0 0175 0 0031 0 9780	0 0285 0 0243 0 9935	0 0111 0 0027 0 9917	0 0281 0 0243 0 9931	0 0088 0 0021 0 9931	0 0290 0 0248 0 9927	0 0091 0 0021 0 9927
Pasighat	0 0484 0 0357 0 9896	0 0218 0 0036 0 9625	0 0806 0 0576 0 9736	0 0341 0 0057 0 9131	0 0421 0 0323 0 9925	0 0178 0 0031 0 9788	0 0578 0 0417 0 9854	0 0229 0 0038 0 9592	0 0755 0 0534 0 9761	0 0246 0 0041 0 9541
Shillong	0 0557 0 0411 0 9899	0 0577 0 0082 0 9368	0 0664 0 0546 0 9781	0 0256 0 0053 0 9654	0 0777 0 0519 0 9875	0 1001 0 0111 0 9205	0 0565 0 0417 0 9900	0 0701 0 0090 0 9364	0 0551 0 0412 0 9900	0 0209 0 0009 0 9372
Silchar	0 0573 0 0462 0 9780	0 0212 0 0045 0 9769	0 0638 0 0491 0 9655	0 0233 0 0057 0 9701	0 0641 0 0514 0 9795	0 0244 0 0049 0 9734	0 0561 0 0452 0 9775	0 0202 0 0045 0 9772	0 0553 0 0441 0 9761	0 0105 0 0012 0 9760
Tezpur	0 0393 0 0278 0 9921	0 0166 0 0029 0 9921	0 0254 0 0195 0 9915	0 0077 0 0015 0 9898	0 0519 0 0362 0 9871	0 0230 0 0038 0 9863	0 0382 0 0272 0 9925	0 0159 0 0028 0 9925	0 0358 0 0260 0 9931	0 0073 0 0007 0 9932

Table A4. 1 Parameters of the GEV, GLD and GPD distributions for different levels of the LH-moments

Region	Distribution	η	ξ	α	k
Guwahati	GEV	0	88.1267	20.8261	-0.1920
		1	87.9562	21.5838	-0.1720
		2	87.6470	22.1440	-0.1600
		3	88.0960	21.5292	-0.1720
		4	89.7331	19.7911	-0.2040
	GLD	0	96.5582	15.3129	-0.3000
		1	96.8293	16.4068	-0.2600
		2	96.5748	17.3894	-0.2320
		3	96.5525	17.3978	-0.2320
		4	97.2865	16.5057	-0.2520
	GPD	0	67.8735	40.0735	0.0800
		1	71.2536	33.7353	-0.0160
		2	73.0476	31.2425	-0.0520
		3	75.5521	28.3871	-0.0920
		4	79.3235	24.7381	-0.1440
Imphal	GEV	0	68.7243	22.0844	-0.0560
		1	68.8098	21.7841	-0.0640
		2	68.4107	22.6281	-0.0440
		3	67.2442	24.4173	-0.0080
		4	65.6347	26.4971	0.0280
	GLD	0	77.4480	15.0803	-0.2040
		1	77.6387	15.7335	-0.1760
		2	77.4063	16.9402	-0.1360
		3	76.6252	18.5626	-0.0920
		4	75.3747	20.4740	-0.0480
	GPD	0	45.2606	49.5017	0.3200
		1	51.0321	37.0584	0.1360
		2	52.6251	34.5660	0.1000
		3	52.1787	35.1621	0.1080
		4	50.8097	36.8118	0.1280
Mohanbari	GEV	0	108.5876	24.7456	-0.4560
		1	111.2747	18.9898	-0.5480
		2	114.4320	15.7224	-0.6000
		3	115.1759	15.2098	-0.6080
		4	114.2330	15.7461	-0.6000
	GLD	0	119.0041	20.9738	-0.5000
		1	119.3230	17.0422	-0.5720
		2	120.8866	14.7289	-0.6120
		3	121.3886	14.2584	-0.6200
		4	120.2340	15.0183	-0.6080
	GPD	0	87.9945	36.7087	-0.3320
		1	98.7051	23.1564	-0.5000
		2	105.1010	17.9485	-0.5720
		3	107.3457	16.5576	-0.5920
		4	106.2944	17.1269	-0.5840

North Lakhimpur	GEV	0	132.4629	32.4967	0.0680
		1	132.5034	32.1487	0.6000
		2	132.8855	31.2634	0.0440
		3	134.7056	28.3465	-0.0040
		4	137.5094	24.8664	-0.0600
	GLD	0	144.7272	20.6659	-0.1280
		1	145.2755	21.8243	-0.0840
		2	145.1911	22.4170	-0.0680
		3	145.5779	21.5438	-0.0880
		4	146.7518	19.8300	-0.1240
	GPD	0	95.0588	83.7516	0.5480
		1	104.6007	60.4774	0.3160
		2	110.2494	50.6321	0.2160
		3	116.9921	41.1781	0.1160
		4	123.6295	33.4831	0.0280
Pasighat	GEV	0	179.4857	52.9934	-0.2280
		1	181.8514	45.3775	-0.3040
		2	184.2614	41.8380	-0.3360
		3	185.5276	40.4495	-0.3480
		4	185.0028	40.9355	-0.3440
	GLD	0	201.2161	39.8309	-0.3240
		1	200.7994	36.7982	-0.3640
		2	201.4011	35.3414	-0.3800
		3	201.6417	34.9575	-0.3840
		4	200.8384	35.8088	-0.3760
	GPD	0	129.0873	98.1612	0.0200
		1	148.6475	64.6766	-0.1920
		2	158.2812	53.5147	-0.2680
		3	162.6944	49.4492	-0.2960
		4	164.2274	48.2522	-0.3040
Shillong	GEV	0	123.1515	46.6927	0.1360
		1	123.2384	46.1865	0.1280
		2	125.4553	40.6763	0.0520
		3	128.9628	35.0935	-0.0200
		4	131.5196	31.9434	-0.0600
	GLD	0	140.5164	28.6303	-0.0840
		1	141.3494	30.2614	-0.0360
		2	141.4395	29.1413	-0.0600
		3	142.4142	26.9184	-0.1000
		4	143.3924	25.4737	-0.1240
	GPD	0	66.6461	131.4254	0.6880
		1	81.8163	91.7894	0.4160
		2	95.8082	66.3952	0.2280
		3	107.1710	50.5080	0.0960
		4	114.0606	42.5995	0.0240
Silchar	GEV	0	127.2845	41.4230	-0.0480
		1	126.8492	43.7746	-0.0120
		2	127.2740	42.8245	-0.0240
		3	129.1599	39.9865	-0.0560
		4	131.6486	37.0200	-0.0880

	GLD	0	143.5520	28.1337	-0.2000
		1	144.4652	30.8505	-0.1360
		2	144.2591	31.7718	-0.1200
		3	144.6023	30.9996	-0.1320
		4	145.4382	29.8397	-0.1480
	GPD	0	83.2407	92.9755	0.3280
		1	90.3560	77.2992	0.2080
		2	97.3223	66.0055	0.1240
		3	104.5976	56.4702	0.0520
		4	111.4662	48.8475	-0.0080

Tezpur	GEV	0	91.2443	23.0373	0.0360
		1	90.8532	26.3135	0.1360
		2	90.7472	26.6363	0.1440
		3	91.4932	25.2839	0.1160
		4	92.5300	23.8201	0.0880
	GLD	0	100.1223	14.9404	-0.1440
		1	101.1811	17.2132	-0.0280
		2	101.0920	18.1910	0.0080
		3	101.0408	18.3236	0.0120
		4	101.2051	18.0289	0.0040
	GPD	0	65.2151	57.4905	0.4920
		1	66.9149	53.2403	0.4360
		2	70.7152	46.0112	0.3480
		3	74.9194	39.4603	0.2680
		4	78.7234	34.3881	0.2040

Table A4.2 RRMSE values for different recurrence intervals of GEV, GPD and GLD distributions

Region	Distribution	η	2	5	10	20	50	100
Imphal	GEV	0	**0.0575**	**0.0652**	**0.0771**	0.1007	0.1434	0.1759
		1	0.0586	0.0681	0.082	0.1002	0.1426	0.1947
		2	0.0609	0.0702	0.0829	0.1011	0.1349	0.1812
		3	0.0662	0.0749	0.0806	**0.0989**	**0.1299**	**0.1713**
		4	0.0776	0.0824	0.0911	0.1048	0.1457	0.1792
	GPD	0	0.0674	**0.0728**	**0.0683**	**0.0748**	**0.0951**	**0.1142**
		1	**0.058**	0.0767	0.0796	0.094	0.1176	0.1463
		2	0.0581	0.0781	0.0885	0.0948	0.124	0.1524
		3	0.0601	0.0792	0.0874	0.0927	0.1207	0.1507
		4	0.0683	0.0814	0.0896	0.096	0.116	0.147
	GLD	0	**0.0546**	**0.0638**	0.0829	0.1137	0.1677	0.2094
		1	0.0582	0.0648	0.0823	0.1062	0.1603	0.2144
		2	0.0613	0.0675	**0.0776**	0.1026	0.1491	0.1989
		3	0.0682	0.0718	0.0799	0.0967	0.1414	0.1868
		4	0.0801	0.0739	0.0795	**0.095**	**0.1325**	**0.1726**
Mohanbari	GEV	0	0.0531	**0.0862**	**0.1277**	**0.1784**	**0.273**	**0.3862**
		1	0.0439	0.0866	0.1287	0.1819	0.2895	0.3945
		2	**0.0348**	0.0874	0.1334	0.1862	0.3171	0.3955
		3	0.037	0.0929	0.1446	0.2013	0.314	0.4021
		4	0.0495	0.0966	0.1561	0.2105	0.3297	0.3884
	GPD	0	0.0547	0.0959	**0.127**	**0.1775**	**0.2648**	**0.3539**

		1	0.0396	0.0963	0.135	0.1894	0.3139	0.3786
		2	**0.0324**	0.0942	0.1356	0.1929	0.317	0.3916
		3	0.0358	0.0978	0.1441	0.2115	0.3065	0.42
		4	0.0461	**0.0931**	0.1555	0.2211	0.3415	0.4083
	GLD	0	0.0516	**0.086**	**0.1227**	**0.1864**	**0.2884**	0.4155
		1	0.0443	0.0862	0.1303	0.1876	0.2919	**0.3957**
		2	**0.0377**	0.0888	0.13	0.1941	0.3027	0.415
		3	0.0382	0.0915	0.1394	0.1961	0.3066	0.4071
		4	0.0518	0.0981	0.1553	0.236	0.3329	0.4018
North Lakhimpur	GEV	0	0.0425	**0.0468**	**0.0527**	**0.0653**	0.0905	**0.1106**
		1	0.0428	0.0488	0.0535	0.0682	**0.0902**	0.1145
		2	0.0433	0.0504	0.0591	0.0673	0.0939	0.1185
		3	0.0399	0.0505	0.0595	0.0691	0.0995	0.1302
		4	**0.0382**	0.0498	0.0604	0.0749	0.1077	0.143
	GPD	0	0.0526	**0.0483**	**0.0448**	**0.0455**	**0.0553**	**0.0672**
		1	0.046	0.0536	0.0529	0.0587	0.0697	0.0875
		2	0.0405	0.0558	0.0591	0.0643	0.081	0.1007
		3	**0.036**	0.0536	0.0627	0.0712	0.0928	0.1158
		4	0.0363	0.0518	0.0616	0.0777	0.1041	0.1349
	GLD	0	**0.04**	0.0463	0.0589	0.0791	0.1082	0.143
		1	0.0411	**0.0448**	0.0547	0.0705	**0.1018**	**0.1316**
		2	0.0436	0.046	**0.054**	**0.0694**	0.103	0.1385
		3	0.0414	0.0485	0.0555	0.0701	0.1036	0.1413
		4	0.041	0.0491	0.0578	0.0723	0.1116	0.1588
Pasighat	GEV	0	0.0579	**0.0759**	**0.0981**	**0.1373**	**0.2118**	**0.2661**
		1	0.0531	0.0787	0.1074	0.1442	0.2249	0.3346
		2	**0.0495**	0.0817	0.1121	0.1512	0.2195	0.3208
		3	0.0499	0.0835	0.1099	0.1581	0.2274	0.3185
		4	0.0556	0.0865	0.115	0.1502	0.2338	0.3148
	GPD	0	0.0637	**0.0873**	**0.096**	**0.1151**	**0.1583**	**0.1971**
		1	0.0502	0.0885	0.1075	0.1467	0.2059	0.2761
		2	0.0458	0.0893	0.1201	0.1486	0.2248	0.2974
		3	**0.0454**	0.089	0.1206	0.1497	0.2243	0.3056
		4	0.0524	0.0902	0.1216	0.1558	0.2285	0.2986
	GLD	0	0.0554	**0.074**	0.1024	**0.1422**	0.2415	**0.3056**
		1	0.052	0.0764	**0.1015**	0.1451	0.2359	0.3123
		2	**0.0495**	0.0803	0.1037	0.1466	**0.2219**	0.3216
		3	0.0525	0.0829	0.1096	0.1503	0.2275	0.3145
		4	0.0571	0.0862	0.1166	0.1523	0.2376	0.3107
Shillong	GEV	0	0.0622	**0.0575**	**0.0628**	0.0727	**0.0921**	**0.1178**
		1	0.064	0.0593	0.0639	**0.0719**	0.0948	0.1178
		2	0.0652	0.0609	0.0693	0.0774	0.1077	0.1333
		3	0.0517	0.0626	0.0717	0.0871	0.1192	0.1556
		4	**0.049**	0.0636	0.0743	0.0892	0.1245	0.1671
	GPD	0	0.0747	**0.0582**	**0.0478**	**0.0464**	**0.0575**	**0.0687**
		1	0.0646	0.0653	0.0618	0.0607	0.0749	0.0872
		2	0.0542	0.0696	0.0708	0.0787	0.093	0.1135
		3	0.0469	0.065	0.0745	0.0858	0.1079	0.139
		4	**0.045**	0.0646	0.07771	0.09	0.1187	0.1562
	GLD	0	0.0556	0.0575	0.0683	0.086	0.1205	0.1466
		1	0.0594	**0.0573**	**0.0635**	0.083	**0.11**	**0.1396**
		2	0.0591	0.0592	0.0698	**0.0824**	0.1183	0.1513
		3	0.0537	0.0597	0.0701	0.0836	0.1246	0.1652

		4	**0.0533**	0.0604	0.0711	0.0886	0.1292	0.1727
Silchar	GEV	0	**0.0574**	**0.066**	0.0774	0.098	0.1396	0.1731
		1	0.0606	0.0686	**0.0755**	0.0972	**0.129**	**0.1647**
		2	0.0616	0.0711	0.0831	**0.0952**	0.1334	0.169
		3	0.0581	0.0716	0.0834	0.0961	0.1375	0.1793
		4	0.0584	0.0717	0.0846	0.1026	0.144	0.1884
	GPD	0	0.0672	**0.0711**	**0.0703**	**0.0764**	**0.093**	**0.1111**
		1	0.0643	0.0758	0.0772	0.0825	0.1089	0.1284
		2	0.0569	0.0769	0.0809	0.092	0.1183	0.1493
		3	0.0543	0.076	0.0862	0.0977	0.1279	0.1642
		4	**0.0526**	0.0741	0.0883	0.1022	0.135	0.1765
	GLD	0	**0.0534**	**0.0644**	0.0839	0.1096	0.1619	0.2227
		1	0.0605	0.0646	**0.076**	0.1042	0.146	0.1925
		2	0.0614	0.0662	0.0774	0.0995	0.1455	**0.1825**
		3	0.0628	0.0685	0.079	**0.098**	**0.1435**	0.1991
		4	0.0617	0.0688	0.0806	0.099	0.1436	0.199
Tezpur	GEV	0	0.1884	**0.0499**	0.0576	0.071	0.099	0.1214
		1	0.0502	**0.0499**	**0.0525**	**0.0618**	**0.0819**	0.1038
		2	0.0499	0.0511	0.0551	0.0621	0.0823	**0.1005**
		3	0.0498	0.0521	0.0547	0.0636	0.0827	0.1037
		4	**0.0492**	0.0532	0.0583	0.0661	0.0868	0.1095
	GPD	0	0.0535	**0.0537**	**0.0485**	**0.0508**	0.0634	0.0736
		1	0.0509	0.0556	0.05	0.0516	**0.0617**	**0.0734**
		2	0.0478	0.0569	0.0541	0.0578	0.067	0.0811
		3	0.0463	0.0566	0.061	0.061	0.0743	0.0888
		4	**0.045**	0.0553	0.0611	0.0631	0.0798	0.0984
	GLD	0	**0.0418**	0.048	0.0637	0.0837	0.12	0.1544
		1	0.04874	**0.0462**	0.0559	0.0679	0.0976	0.1202
		2	0.0501	0.0492	**0.0537**	**0.0645**	**0.0892**	**0.1144**
		3	0.0503	0.0508	0.0558	0.0652	0.0903	0.1163
		4	0.0503	0.0507	0.0569	0.0663	0.0916	0.1174

Table A4. 3 RBIAS values for different recurrence intervals of GEV, GPD and GLD distributions

Region	Distribution	η	2	5	10	20	50	100
Imphal	GEV	0	0.0028	-0.0038	-0.0029	-0.0033	**0.0003**	**-0.0016**
		1	0.003	**-0.0029**	**-0.00086**	**-0.0029**	0.0021	0.0053
		2	0.0018	0.00038	-0.0031	-0.0036	-0.0097	0.0093
		3	0.0021	-0.0019	-0.0038	-0.0061	-0.0037	-0.0023
		4	**0.0017**	0.0041	-0.0027	-0.0093	-0.0093	-0.0097
	GPD	0	0.0023	**0.00007**	**-0.0005**	**-0.0054**	0.0062	0.0108
		1	0.0024	-0.0027	-0.0033	-0.0055	**0.0032**	0.007
		2	0.0011	0.0015	-0.0059	-0.0076	-0.0041	0.0007
		3	0.0022	0.002	-0.0011	-0.0054	-0.0036	**-0.0003**
		4	**-0.0006**	0.0019	-0.0044	-0.0062	-0.004	-0.0065
	GLD	0	0.0025	-0.0026	-0.007	**0.0015**	**0.0019**	0.0049
		1	0.0055	-0.0039	-0.0057	-0.006	-0.0055	0.0058
		2	0.0055	**-0.0014**	-0.0043	-0.0075	-0.0107	**-0.001**
		3	**0.0015**	0.0022	**-0.0011**	-0.0118	-0.0052	-0.0045
		4	0.0029	0.0035	-0.0062	-0.0072	-0.0089	0.0037
Mohanbari	GEV	0	0.007	-0.0052	-0.0156	**-0.0278**	**-0.0594**	**-0.0445**

		1	0.0074	**0.0052**	-0.0124	-0.0482	-0.0779	-0.1044
		2	**0.0031**	0.0096	-0.0117	-0.0372	-0.0967	-0.1245
		3	-0.0042	0.0155	**-0.001**	-0.0336	-0.0966	-0.1374
		4	-0.0114	0.0197	0.0132	-0.0345	-0.095	-0.1425
	GPD	0	0.0063	**-0.0044**	-0.0127	**-0.023**	**-0.0393**	-0.0319
		1	0.0049	0.0084	-0.0115	-0.0322	-0.0643	-0.1052
		2	**0.0025**	0.0124	-0.0098	-0.0297	-0.0803	**-0.01243**
		3	-0.0032	0.0159	**-0.0014**	-0.0332	-0.0865	-0.0341
		4	-0.0142	0.0183	0.0067	-0.0289	-0.0813	-0.1257
	GLD	0	0.0055	-0.0111	-0.0176	-0.0394	**-0.0454**	**-0.0308**
		1	0.009	**0.0035**	-0.0176	-0.0477	-0.0816	-0.111
		2	0.0059	0.0128	-0.0087	-0.0364	-0.0968	-0.137
		3	**-0.0021**	0.0141	**-0.0045**	-0.0372	-0.0972	-0.1224
		4	-0.0125	0.0189	0.0093	**-0.0229**	-0.0822	-0.1329
North Lakhimpur	GEV	0	0.0011	**-0.0002**	**-0.0002**	-0.0043	0.0044	0.0069
		1	0.0016	-0.0022	-0.0021	**-0.0032**	0.0031	0.0053
		2	0.0011	0.0006	-0.004	-0.0049	**-0.0024**	0.0005
		3	**0.0003**	0.0011	-0.001	-0.0043	-0.0033	**-0.0003**
		4	-0.0011	0.0013	-0.0017	-0.0072	-0.0067	-0.0021
	GPD	0	0.0018	**0.0003**	-0.0038	**0.0004**	0.0031	0.0049
		1	0.0018	-0.0007	-0.0026	-0.0037	0.0017	0.0034
		2	-0.0011	0.0011	-0.0038	-0.0017	**-0.0013**	0.0021
		3	**-0.0004**	0.0008	**-0.0017**	-0.002	-0.0017	**0.0012**
		4	-0.0023	0.002	-0.0023	-0.0068	-0.0085	-0.0038
	GLD	0	0.0001	-0.0018	-0.003	**-0.0008**	**-0.0012**	0.0057
		1	0.0017	-0.0028	-0.004	-0.0045	-0.0036	**-0.0008**
		2	0.0004	**0.00009**	-0.0045	-0.0032	-0.004	0.0016
		3	**-0.0001**	-0.0005	-0.0017	-0.0047	-0.0043	-0.0028
		4	-0.0006	0.0031	**-0.0013**	-0.0092	-0.009	-0.0018
Pasighat	GEV	0	0.0043	-0.0053	-0.0067	**-0.0093**	**-0.0066**	**-0.0098**
		1	0.0052	**-0.0031**	-0.0048	-0.0135	-0.014	-0.0108
		2	0.0033	0.0034	-0.0064	-0.0156	-0.0386	-0.0269
		3	**-0.0005**	0.0063	-0.011	-0.0184	-0.0336	-0.0423
		4	-0.003	0.0097	**-0.0019**	-0.021	-0.0384	-0.0548
	GPD	0	0.003	**0.00009**	-0.0022	**-0.0111**	**0.0044**	0.0095
		1	0.0036	-0.002	-0.0072	-0.0144	-0.0089	**-0.0064**
		2	**0.0017**	0.0048	-0.0087	-0.021	-0.028	-0.032
		3	-0.0018	0.0072	**-0.001**	-0.0178	-0.0318	-0.0342
		4	-0.0059	0.0098	0.0027	-0.0243	-0.029	-0.047
	GLD	0	**0.0015**	-0.0058	-0.0109	**-0.0095**	**-0.0046**	**0.0038**
		1	0.0071	**-0.0015**	-0.0149	-0.02	-0.0241	-0.0317
		2	0.0031	0.0033	-0.0113	-0.0164	-0.0381	-0.0421
		3	0.0021	0.0037	-0.0025	-0.0193	-0.0416	-0.0489
		4	-0.0042	0.0056	**-0.0015**	-0.0222	-0.0382	-0.0493
Shillong	GEV	0	**0.0009**	-0.002	-0.0006	**0.0012**	-0.0035	0.0072
		1	0.001	**0.0001**	-0.0003	-0.0041	**-0.0011**	0.0026
		2	0.0012	-0.0014	-0.0054	-0.0015	-0.0041	0.0047
		3	0.0011	-0.0006	-0.0011	-0.0054	-0.0069	**0.0022**
		4	-0.0026	0.0027	**0.0001**	-0.0047	-0.0061	-0.0056
	GPD	0	0.0006	-0.0012	**-0.0004**	0.0035	0.0044	0.0129
		1	-0.0003	**0.0002**	-0.0045	**-0.002**	**-0.0001**	0.0088
		2	0.0006	-0.0015	-0.0013	-0.0055	-0.0025	0.0032
		3	0.0001	0.0018	-0.0014	-0.0036	-0.0038	**0.0011**

		4	**-0.0001**	0.0017	-0.0024	-0.0057	-0.0101	-0.0023
	GLD	0	0.0011	-0.0011	**-0.0005**	-0.0054	0.0059	0.0083
		1	0.0023	-0.0031	-0.0027	**-0.0034**	**0.0034**	0.0062
		2	0.0021	**0.0004**	-0.0051	-0.0061	-0.0034	**-0.0006**
		3	0.0012	0.0013	-0.0017	-0.006	-0.005	-0.0017
		4	**-0.0009**	0.0015	-0.0025	-0.0089	-0.0086	-0.0035
Silchar	GEV	0	0.002	**-0.0006**	**-0.0012**	-0.0076	0.0052	0.0085
		1	0.0027	-0.003	-0.0035	**-0.0052**	**0.003**	0.0061
		2	-0.0018	0.001	-0.0059	-0.0078	-0.0049	**-0.0014**
		3	**0.0004**	0.002	-0.0015	-0.0068	-0.0061	-0.0022
		4	-0.0019	0.0021	-0.0024	-0.0102	-0.01	-0.0043
	GPD	0	**0.00001**	-0.0064	-0.0023	-0.0055	0.0032	0.0079
		1	-0.0005	-0.0039	-0.0038	-0.006	**0.0028**	0.0101
		2	0.0021	**0.0005**	-0.0049	**-0.005**	-0.0046	0.0069
		3	0.0009	0.0033	**-0.002**	-0.0085	-0.0089	-0.0053
		4	-0.0041	0.0007	-0.0034	-0.0063	-0.0095	**0.0018**
	GLD	0	**-0.0014**	-0.0026	-0.0078	-0.0034	**0.0002**	0.0072
		1	0.0015	-0.0023	-0.0065	**-0.0007**	0.0018	0.009
		2	0.0054	**-0.0013**	-0.0041	-0.0065	-0.0009	-0.0067
		3	0.0019	0.0017	-0.0064	-0.0094	-0.0134	**-0.002**
		4	**-0.0014**	0.0018	**-0.0022**	-0.0081	-0.0091	-0.0139
Tezpur	GEV	0	-0.0043	**-0.0008**	-0.0015	-0.0018	0.0038	**0.0002**
		1	0.0008	-0.0019	**-0.0006**	-0.004	**0.0003**	0.0026
		2	0.0032	-0.0012	-0.0033	**-0.0014**	-0.0017	0.0019
		3	-0.0007	-0.0011	-0.0025	-0.0018	-0.0013	0.0024
		4	**-0.0001**	0.0031	-0.0034	-0.0037	-0.0026	-0.0019
	GPD	0	0.0028	-0.0024	-0.0035	**0.0004**	0.0051	0.0074
		1	0.002	**0.0006**	**-0.0001**	-0.0039	0.004	0.0065
		2	0.0015	-0.0017	-0.0018	-0.0033	0.0028	0.0053
		3	**0.00004**	0.0014	-0.0037	-0.0044	**-0.0018**	0.0021
		4	-0.0014	0.0017	-0.0003	-0.0033	-0.0022	**-0.0009**
	GLD	0	0.0007	-0.0038	-0.0019	**-0.0005**	0.0026	0.0085
		1	**0.0003**	**0.00009**	-0.0043	-0.0022	0.0021	0.0021
		2	0.0031	0.0005	-0.0011	-0.0023	-0.0027	0.0047
		3	0.0019	-0.0006	**-0.0003**	-0.005	**-0.0012**	0.0002
		4	0.0004	0.0011	-0.0016	-0.0021	-0.002	**0.0001**