Uncertainty Analysis of Regional Rainfall Frequency Estimates in Northeast India

Nilotpal Debbarma [1*], Parthasarathi Choudhury [2], Parthajit Roy [3], Shivam Agarwal [1]

[1] *PhD Scholar, Civil Engineering Department, NIT Silchar, Assam, India.*

[2] *Professor, Civil Engineering Department, NIT Silchar, Assam, India.*

[3] *Associate Professor, Civil Engineering Department, NIT Silchar, Assam, India.*

Abstract

Estimation of rainfall quantile is an important step in regional frequency analysis for planning and design of any water resources project. Related evaluations of accuracy and uncertainty help to further assist in enhancing the reliability of design estimates. In this study, therefore, we investigate the accuracy and uncertainty of regional frequency analysis of extreme rainfall computed from genetic algorithm-based clustering. Uncertainty assessment is explored with prediction of quantiles with a new spatial Information Transfer Index (ITI) and Monte Carlo simulation framework. And, accuracy assessment is done with the comparison of regional growth curves to at-site analysis for each homogenous region. Further, uncertainty assessment with the ITI method is compared with Maximum Likelihood Estimation (MLE) optimized by a Genetic Algorithm (GA) to check the suitability of the method. Results obtained suggest the ITI-based uncertainty assessment for regional estimates outperformed those of at-site estimates. The MLE-GA method based on at-site estimates was found to be better than at-site estimates based on L-moments, suggesting the former as a better alternative to compare with regional frequency estimates. Moreover, minimal bias and least deviation of the regional growth curve were obtained in the rainfall regions. The confidence intervals of regional estimates were seen to be well within the bounds of normality assumptions.

Keywords: L-Moments; Monte Carlo; Information Transfer Index; MLE; GA; Rainfall.

1. Introduction

The Brahmaputra and Barak basins in northeastern India are among the country's most disaster-prone locations, with severe rainstorms and cloud bursts occurring annually during the monsoon season. The frequency of extremely heavy downpours in the basins has been shown to fluctuate widely over the region. As a result, human life and property are seriously damaged, which has an impact on the region's total socioeconomic activity. With a limited network of rain gauge stations, flood related information and mitigation measures have always been insufficient in the region. Methods such as Regional Frequency Analysis (RFA) have been frequently employed in such situations that transfer data from gauged locations to places with little or no data [1, 2]. Several studies can be found with the application of the RFA technique in the region [3-7]. Regional frequency analysis of extreme rainfall for any region aims to provide a detailed description of the distribution of rainfall events and predict probable estimates for a given return time. However, the

application of the RFA technique is always associated with some amount of uncertainty. Quality of data, number of available records, and processes involved in any regional frequency analysis (RFA) are important sources of uncertainty, and so is crucial to analyse the uncertainty inherent in their applications. The fitting of the probability distribution and parameter estimation is an important factor affecting RFA analysis [8, 9]. With uncertainty and suitability involved in the selection of an appropriate probability distribution for any regional frequency analysis, it is crucial to have a performance comparison of fitted distributions against at-site analysis for the region. Numerous works on uncertainty analysis in regional frequency analysis of rainfall [10-13] have been published. Notable recent studies on the comparison of uncertainty analysis of rainfall from regional and at-site analysis can be found in the studies of [14-16], but very scarce studies are available for north east India. The available studies are limited to only selection of probability distribution and estimation of quantiles [3, 5]. Resulting quantiles from RFA with small sample sizes are speculative, and constructing confidence intervals has been the simplest and most widely used approach to evaluate the uncertainty. There is rarely any comparison study of uncertainties associated with at-site and regional analysis in the study region, which is a necessary area to be explored. The study will give an assessment of the extreme rainfall behaviour with respect to fitted distributions both at regional and at-site levels.

Besides evaluating confidence intervals based on Monte Carlo simulation, several other approaches for analysing uncertainty have been developed, including generalised likelihood uncertainty estimation (GLUE), the Bayesian method, and Markov chain Monte Carlo (MCMC), among others [17-21]. The majority of them were used to quantify uncertainty in hydrological models. But the approach for analysing uncertainty for at-site and regional frequency analysis using the Monte Carlo simulation and entropy-based information transfer index (ITI) framework till date has not been explored. In this study, the uncertainty is performed with generation of new samples using Monte Carlo simulation from entropy dependent weights at an unmeasured site for rainfall estimates. Compared to other uncertainty methods, there is rarely any study of the present approach in frequency analysis. The entropy concept has been used in many hydrological studies [22-25], but its application in uncertainty analysis in RFA has not been done. The proposed method is based on the idea that information at a new station is more accurately generated from nearby stations when stations with a higher amount of shared information are selected rather than stations based on proximity or distance.

The majority of the regional frequency analysis studies conducted in India's northeast area concentrated on the Brahmaputra basin or important chosen stations from the whole northeast region. However, relatively little research on regional frequency studies of yearly extreme rainfall, including rain gauge stations from the Barak basin, is available. During the monsoons, the Barak basin is severely prone to flooding, producing flood issues comparable to those seen in the Brahmaputra basin. So, with the inclusion of stations from the Barak basin in the study, will provide a more comprehensive and enlarged perspective of the extreme rainfall scenario in the northeast area. Moreover, to the best of our knowledge, research on the uncertainty of regional rainfall quantiles for homogenous rainfall regions in the Brahmaputra and Barak basins is scarce. The study therefore aims to assess the uncertainty of extreme rainfall estimates derived from regional frequency analysis and to perform a comparison with at-site analysis. To assess the suitability of design quantiles, two approaches are investigated: (i) uncertainty estimation using coefficient of variation of rainfall estimates and the development of confidence intervals; and (ii) using the framework of ITI and Monte Carlo simulation and comparing it to at-site frequency. Furthermore, to investigate the applicability of ITI-based weight determination with different parameter estimation methods, MLE estimation optimised by a genetic algorithm (GA) is investigated.

2. Study Area and Rainfall Data

The study region in northeast India concentrates on the southern part of the Brahmaputra basin and the Barak basin. The Barak basin, just beside the Brahmaputra basin, has active floodplains with vast marshy regions that are annually inundated by severe floods. The altitudinal pattern in the north east changes drastically from location to location, resulting in erratic rainfall occurrences. During the monsoon season, heavy rains and cloud bursts are common, wreaking havoc on the region and causing widespread landslides and erosion. Amount of precipitation is particularly unpredictable in the region, making future rainfall scenarios highly vulnerable. For the study, annual maximum daily rainfall data for 33 stations is collected from the Regional Meteorological Centre, Guwahati, for a period of 20 years, and their locations in the research region are given in Figure 1. The list of stations included in the study are Silchar (1), Dholai (2), Goalpara (3), Guwahati (4), North Lakimpur (5), Choudhoughat (6), Batadighat (7), Kampur (8), Sibsagar (9), Beki Rd. Bridge (10), Dibrugarh (11), Jorhat (12), Neamatighat (13), Kherunighat (14), Bokajan (15), Gossaigaon (16), Kokrajhar (17), Tezpur (18), Mellabazar (19) Aie NH.Xing(20), Dharamtul(21), Golaghat(22), Dhollabazar (23), Margherita (24), Gharmura (25), Shillong (26), Cherrapunjee (27), Mawsynram (28), Kohima (29), Imphal (30), Aizwal (31), Agartala (32), and Kailashahar (33).
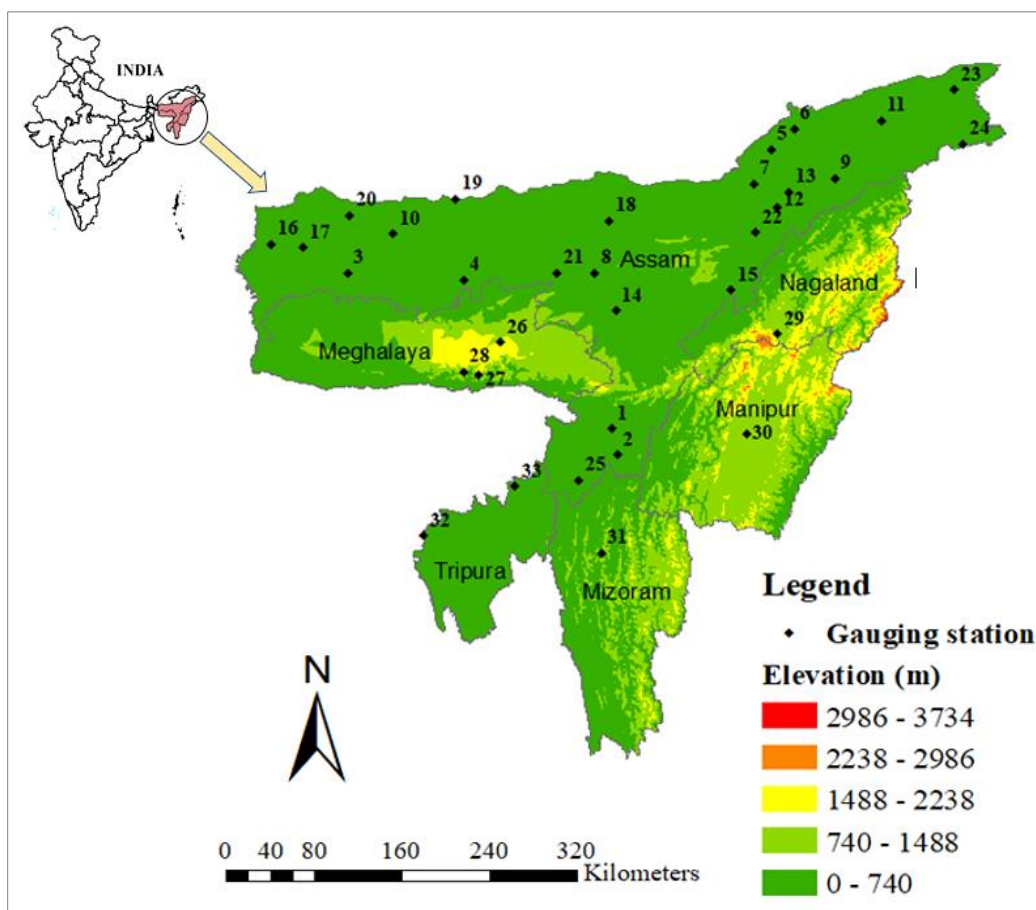
Figure 1. Location of gauging stations in study area of Brahmaputra and Barak basin

3. Research Methodology

3.1. Formation of Homogenous Regions and Heterogeneity Measurements

The study used genetic algorithm-based clustering to designate homogeneous rainfall areas, with the Davies-Bouldin index as the fitness function. Based on the multi-criteria decision technique and heterogeneity measures proposed in [2], three optimal station groups were determined and found to be homogeneous. Clustering was done using seven station characteristics: latitude, longitude, altitude, annual daily maximum average, greatest annual daily maximum, lowest annual daily maximum, and annual maximum series coefficient of variation. Prior to clustering, the variables were standardised using the max-min transformation. Verification of the homogeneity of identified homogenous regions in regional frequency analysis is very important and is done in the present study using the heterogeneity measure [2]. According to Tasker et al. (1998) [2], a region is declared acceptably homogenous when H<1, a possibly heterogenous when 1≤H<2 and definitely heterogenous if H ≥ 2. More information about the procedure can be obtained from [2]; And, the study framework for comparing regional and at-site precipitation frequency analyses is summarised in Figure 2, which addresses the steps followed in determining uncertainty.

3.2. Choosing Best Fit Distribution and Accuracy of Quantiles

The best-fit probability distribution for all three regions was determined by testing five three-parameter candidate distributions: generalised normal (GNO), generalised Pareto (GPA), generalised extreme value (GEV), generalised logistic (GLO), and Pearson type 3 (PE3). The goodness-of-fit metric as suggested in [2] for homogenous regions is considered to find the best distribution and is calculated as

$$\left|Z^{dist}\right| = (\tau_4^{dist}\beta - \tau_4^R + \beta_4)/\sigma_4 \tag{1}$$

where "dist" is the candidate distribution; τ_4 the regional average L-kurtosis value calculated in simulation; β_4 and σ_4 the bias and standard deviation respectively of regional average L-kurtosis (τ_4^R) of Monte Carlo simulation samples performed by Kappa distribution. For all values of $\left|Z^{dist}\right| \leq 1.64$, the corresponding candidate distributions are considered fit and acceptable at 90 % confidence level. And to decide on the best-fit distribution, the candidate distribution with lowest $\left|Z^{dist}\right|$ is selected as the best distribution for the region. The annual maximum rainfall quantiles for different probabilities of non-exceedance F are then calculated with selected fitted distributions using the method of index flood approach.
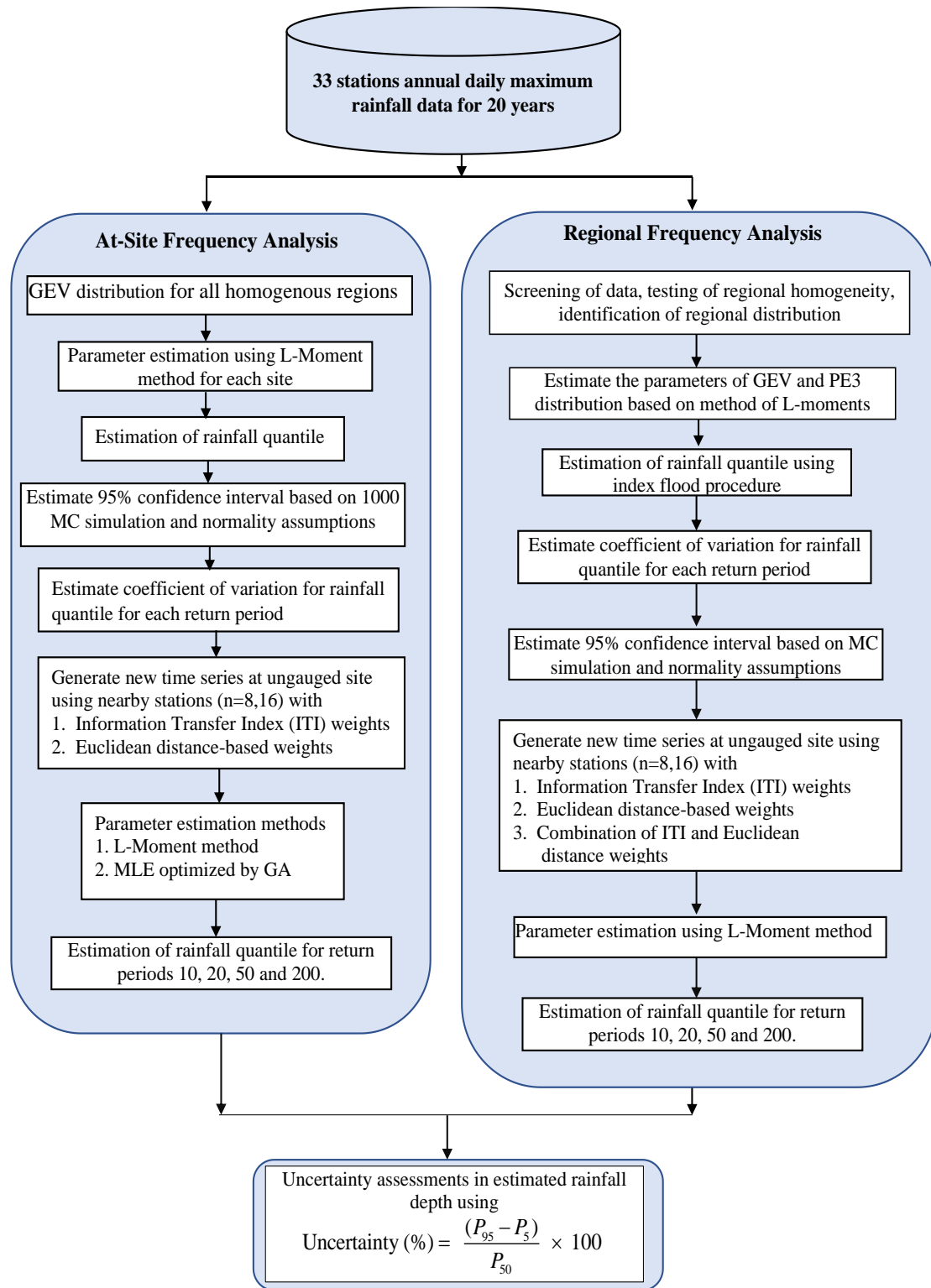
Figure 2. An overview of study framework for estimation of uncertainty

To evaluate robustness of the regional quantiles for each region, the procedure mentioned by Tasker et al. (1998) [2] involving generation of regional average L-moments from Monte Carlo simulations is used. The simulation involves generation of quantile estimates for various return periods, and at a given m^{th} repetition, the estimated quantiles for a given non-exceedance probability F, $\hat{Q}_i^{[m]}(F)$ is estimated and compared with true values of $Q_i(F)$. The relative error of this estimate at a given site i and for a non-exceedance probability F is expressed as;

$$\frac{\hat{Q}_i^{[m]}(F) - \hat{Q}_i(F)}{\hat{Q}_i(F)} \tag{2}$$

This quantity is squared and averaged for M repetitions to obtain the relative bias and mean relative quadratic error as;

$$B_i(F) = M^{-1} \sum_{m=1}^{M} \frac{\hat{Q}_i^{[m]}(F) - \hat{Q}_i(F)}{\hat{Q}_i(F)} \tag{3}$$

$$R_i(F) = \left[M^{-1} \sum_{m=1}^{M} \left\{ \frac{\hat{Q}_i^{[m]}(F) - \hat{Q}_i(F)}{\hat{Q}_i(F)} \right\}^2 \right]^{1/2} \tag{4}$$

And the summary performance of all stations in a region, is expressed by regional relative bias and relative root mean square error as;

$$B^R(F) = N^{-1} \sum_{i=1}^{N} B_i(F) \tag{5}$$

$$A^R(F) = N^{-1} \sum_{i=1}^{N} |B_i(F)| \tag{6}$$

$$R^R(F) = N^{-1} \sum_{i=1}^{N} R_i(F) \tag{7}$$

3.3. Uncertainty Analysis of Fitted Distribution Parameters and Quantiles

For higher return periods, the quantiles computed have a higher degree of uncertainty. Uncertainty in parameter estimation and its stability are two important types of uncertainty associated with any given quantile prediction. These uncertainties will be examined in the present study for both regional and at-site analysis with the help of a Monte Carlo simulation framework. Furthermore, confidence intervals are developed to compare accuracy of computed quantiles both from at-site and regional analysis. In the present study, coefficient of variation is used as a measure of uncertainty to test the parameter stability of identified distributions. A C_v for a particular quantile is defined as $C_v = \sigma/\mu$ where, σ and μ are the standard deviation and mean of quantiles estimated from various GEV and PE3 distributions. With 1000 Monte Carlo simulations, random distinct sample sets is generated each time, having the same distribution with different parameters.

For the comparison of uncertainty in prediction of regional and at-site analysis in each region, a quantitative indicator is proposed that takes into account the spread of confidence intervals. For a given quantile prediction, an average relative width [26, 27] is used and is given as;

$$ARW = \frac{1}{n} \sum \frac{Limit_{upper} - Limit_{lower}}{q(F)} \tag{8}$$

where $Limit_{upper}$ and $Limit_{lower}$ are upper and lower limit of corresponding 95% error bounds, n is the number of stations in a region, q(F) is the estimated quantile for probability of non-exceedance, F. Smaller value of ARW indicates a smaller uncertainty of the estimated quantile.

The uncertainty of predicted quantiles for different return periods is assessed by constructing confidence intervals. The steps for constructing confidence intervals for each return period of a station are as follows:

- First, the parameters for fitted distribution of each region is determined using method of L-moments based on observed data of each site.

- A set of generated data having the same sample size as of the site *i* is obtained.

- Monte Carlo simulation is then carried out and parameters of the fitted distribution for each generated sample is calculated and the precipitation quantiles is estimated.

- The 2.5 and 97.5 percentile values for 95 percent lower and upper bounds are obtained for each return period from each Monte Carlo simulation. Similarly, 95 percent confidence interval boundaries are generated for all other return periods of interest.

The confidence intervals from normality assumption is also constructed for comparision purpose and for a target return period with sample mean (μT) and standard deviation (σT), the upper and lower bounds at 95% confidence interval is calculated.

3.4. Uncertainty Analysis of Parameter Estimation using Entropy based Information Transfer Index

The uncertainty of the parameters identified for probability distributions of clustered regions is assessed using Monte Carlo Simulation and entropy dependent weights. The stability of the identified parameters is studied to assess its capability in identifying extreme rainfall depth at any ungauged site with the help of nearby surrounding sites. The current study differs from previous techniques in [11, 28], in the application of an information transfer index based on entropy to generate new time series for an unmeasured site and the weights proposed herein the study is given as;

$$w_i = \frac{ITI_i}{\sum_{i=1}^{k} ITI_i} \qquad (i=1,2,3,\ldots,k) \tag{9}$$

where, w_i is the extraction entropy weight from station i, k is the number of sites. The extraction entropy weight depends on information transfer index (ITI) as expressed by [22] and is given as;

$$ITI(A,B) = T(A,B) \big/ H(A,B) \tag{10}$$

$$T(A,B) = H(A) + H(B) - H(A,B) \tag{11}$$

where; H(A) and H(B) denote the marginal entropies of rain gauge stations A and B, respectively, whereas H(A,B) denotes their combined entropy. ITI is a symmetric index that quantifies the exchange of information between two stations. The weight value is between 0 and 1, and a greater value suggests a more effective communication of information. The new time series at the ungauged site A is generated with;

$$Y_A = \sum_{i=1}^{k} w_i \cdot Y_i \qquad (i=1,2,3,...,k) \tag{12}$$

To assess the performance of the new extraction entropy weight based on ITI it is compared to two other weighting methods viz. one method based on Euclidean distance [29] and the other method is a combination of ITI and Euclidean distance.

$$w_i = \frac{dist_i^{-1}}{\sum_i^k dist_i^{-1}} \qquad (i=1,2,3,.....,k) \tag{13}$$

$$w_i = \frac{dist_i^{-1} ITI_i}{\sum_i^k dist_i^{-1} ITI_i} \qquad (i=1,2,3,.....,k) \tag{14}$$

New time series samples equal to the number of data length at each site for M extraction sites are generated using Monte Carlo simulation using the parameters of the fitted distribution for the site. The new time series samples at an ungauged Site A can then be generated using Equation 3. For regional analysis, new time series of all sites in the homogeneous region is simulated and taken together. The application of ITI approach in assessing the parameter stability of distributions is also extended to application of Maximum Likelihood Estimation (MLE) optimized by genetic algorithm. The parameter estimation of distributions of individual sites for use in ITI is done by MLE method and optimized by genetic algorithm. And the comparison is done to weighting method based on Euclidean distance. The comparison will help us to understand the applicability of ITI method in determining spatial weights for prediction at ungauged sites. The uncertainty in estimated design rainfall depth is computed as:

$$\text{Uncertainty (\%)} = \frac{(P_{95} - P_5)}{P_{50}} \times 100 \tag{15}$$

where P_{95}, P_5, and P_{50} are the expected design rainfall depths at the 95th, 5th, and 50th percentiles, respectively. The estimates by the ITI based L-Moment estimates, comparison with MLE to fitting GEV and PE3 distributions is explored. Maximum Likelihood Estimation (MLE) is a frequently used technique for estimating parameters of probability distributions, in which the parameter estimates generate the highest chance of occurrence for observations. With numerous application in extreme value models [12, 30], it is considered in the present study for comparison with estimates of ITI based L-Moment estimates. GA was used for optimizing the MLE parameters to arrive at the likelihood of the real value. GA are population-based algorithms and have successfully provided near-real value solutions in various complex problems. The log-likelihood function of the three parameters of the GEV and PE3 distribution is given as;

$$\ln[L(x|\theta_1, \theta_2, \theta_3) = \sum_{i=1}^{n} f(x_i|(\theta_1, \theta_2, \theta_3) \tag{16}$$

where, $f(x|\theta_1, \theta_2, \theta_3)$ is the pdf of GEV or PE3 distribution and $x_i = (x_1, ..., x_n)$ are the observations.

The values of the parameters are then obtained by partially differentiating the log-likelihood function with respect to each parameter and equating it to zero. To further assess the performance of ITI, distance-based and MLE-GA estimates for both regional and at-site.

4. Results and Discussion

The rain gauge stations were tested for trend and randomness of data series using the Mann-Kendall and Ljung Box tests [31]. The study results indicate no trend, and the data were serially independent, making them acceptable for statistical frequency analysis and fitting of probability distributions. The grouping of gauge stations using genetic algorithm based clustering and Euclidean distance measure resulted in three regions. Nine cluster validation measures and MCDM analysis gave three homogenous regions and is given in Table 1. Heterogenity test as proposed by [2] was applied and the final homogenous regions I , II and III composed of 9, 2 and 20 stations respectively after removal of

two discordant stations Golaghat and Goalpara. Applying the goodness of-fit criterion, the fitted frequency distributions for the three regions are selected and presented in Table 1.

Table 1. Homogenous regions obtained from genetic algorithm-based clustering

Regions	No. of Stations	Stations	Heterogeneity Measure			Distribution function
			H1	H2	H3	
I	9	North Lakhimpur, Choudhoughat, Bokajan, Sibsagar, Margherita, Dibrugarh, Jorhat, Dhollabazar, Neamatighat	0.33	0.92	0.37	GEV
II	2	Cherrapunjee, Mawsynram	-0.54	-0.83	-0.15	GEV
III	20	Silchar, Dholai, Guwahati, Kampur, Kherunighat, Imphal, Batadighat, Gossaigaon, Kokrajhar, Tezpur, Mellabazar, Aie N. H. Xing, Dharamtul, Gharmura, Beki Road Bridge, Shillong, Kohima, Aizwal, Agartala, Kailashahar	0.06	-1.11	-1.83	PE3

4.1. Estimation of Precision of Regional Quantile Estimate

In this section, the precision of dimensionless regional growth curve q(F) for each homogenous region is calculated and shown in Table 3. With 10,000 simulations, Monte Carlo simulations procedure was carried out with the selected distribution for each region. In the process, the simulated regional quantiles were compared to the real data for all non-exceedance probability to obtain the precision measurements. Regional relative root mean square error $R^R(F)$ computations show that region III has the least deviation in regional growth curve with slightly higher values obtained in region II. This is an important criterion as it signifies the overall deviation of difference between computed quantiles and true quantiles of all stations in a region. As region II comprises of only two stations and the annual maximum rainfall average are comparatively very high, the bias in estimated regional growth curve may be affected by sampling of a smaller number of stations. $B^R(F)$ and $A^R(F)$ also suggests minimal difference between simulated and true quantiles for all return periods in all the three regions. $A^R(F)$ which gives a measure of bias of estimates of quantiles to be consistently high at some stations and low at others [32] is found to be least in region I, and hence the accuracy of quantile estimates in this region will be better among the three regions. Overall, from the analysis, the estimated regional quantile growth curve in each region is found to be satisfactorily follow the frequency distribution behaviour of all clustered stations in each region.

Table 2. Summary of accuracy of distribution functions in the three homogenous regions

Region	Dist.	F =	0.5	0.8	0.9	0.95	0.98	0.99	0.995	0.998	0.999
		RT (years) =	2	5	10	20	50	100	200	500	1000
I	GEV	q(F)	0.979	1.203	1.332	1.444	1.573	1.659	1.737	1.828	1.889
		$B^R(F)$	0.000	0.000	0.001	0.001	0.003	0.004	0.005	0.007	0.009
		$A^R(F)$	0.044	0.050	0.056	0.063	0.072	0.079	0.085	0.094	0.100
		$R^R(F)$	0.019	0.021	0.024	0.026	0.030	0.033	0.035	0.039	0.042
II	GEV	q(F)	0.990	1.217	1.339	1.439	1.545	1.612	1.669	1.732	1.771
		$B^R(F)$	0.002	0.000	0.001	0.000	0.003	0.006	0.010	0.016	0.022
		$A^R(F)$	0.048	0.050	0.054	0.060	0.072	0.082	0.093	0.108	0.120
		$R^R(F)$	0.043	0.044	0.048	0.054	0.064	0.073	0.084	0.099	0.112
III	PE3	q(F)	0.960	1.234	1.401	1.552	1.736	1.868	1.994	2.156	2.275
		$B^R(F)$	0.002	0.002	0.002	0.002	0.003	0.003	0.004	0.004	0.004
		$A^R(F)$	0.054	0.060	0.069	0.077	0.086	0.093	0.098	0.104	0.108
		$R^R(F)$	0.015	0.017	0.019	0.022	0.024	0.026	0.027	0.029	0.030

4.2. Estimation of Uncertainty Analysis of Regional Quantiles

The Monte Carlo simulation procedure for determining the rainfall amounts for different return periods for each site in the regions were estimated and the measure for uncertainty was expressed using coefficient of variation. The comparison plot of coefficient of variation of regional and at-site analysis results are given in Figures 3 and 4. The GEV distribution was selected as at-site frequency distribution for the whole region in accordance with the studies in [6] based on L-Moments for yearly extreme rainfall. The standard deviation and mean of the quantiles of 1000 simulations for each return period is calculated and the parameter stability of the distributions is assessed using C_v. For return period 10 and 20 years as seen in Figure 3, coefficient of variation is seen to be nearly at same value for regional estimates in all

regions I, II and III, while at-site estimated quantiles are seen to fluctuate with a relatively higher dispersion in regions I and III. The sample size generated in each iteration is equal to the original sample size of 20 for each site. The at-site quantiles from Figure 3 shows that some sites in the at-site analysis have produced better estimates than regional analysis with a lower C_v, while the others were less accurate. For region I, despite both at-site and regional fit GEV distribution, parameters of regional GEV distribution seem to be more stable and reliable. The desirable outcome is to have C_v of all regional estimates lower than at-site, and the probable reason may be due to small sample size of 20 considered in the study. But overall, the Figures 3 and 4 are suggestive of the fact that rainfall depths estimated on regional analysis produce more accurate and reliable estimations. The estimations in regional analysis for region II is seen to be quite similar to at-site estimations for all return periods. The region comprises of only two stations and so the prediction accuracy may not be fully represented for the region. The coefficient of variation is also seen to increase with return period for all regions for both regional and at-site estimations, thereby suggesting increase in uncertainty. But the at-site estimations increase at a higher rate. For example, region I at-site lowest and highest C_v values increased from 0.039 and 0.109 to 0.085 and 0.365 respectively; whereas the increase for corresponding regional values are 0.064 and 0.070 to 0.162 and 0.171 respectively. Comparatively, region III gave very large variations for at-site estimations in all return periods. The results thus indicates that the regional estimate of quantiles is much more reliable and accurate compared to the at-site estimations.
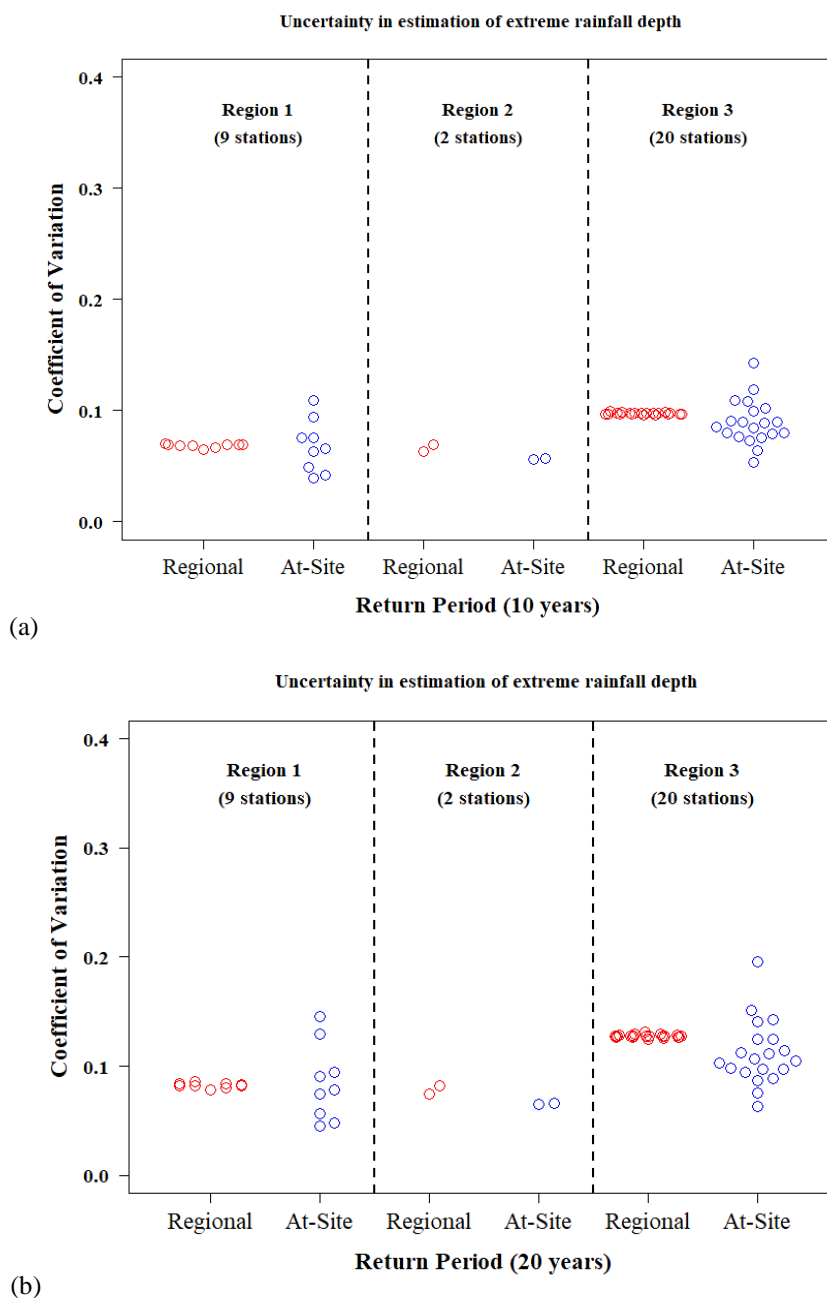

(a)

(b)

Figure 3. Uncertainty of extreme rainfall estimations in homogenous rainfall regions for (a) 10 and (b) 20 years return period
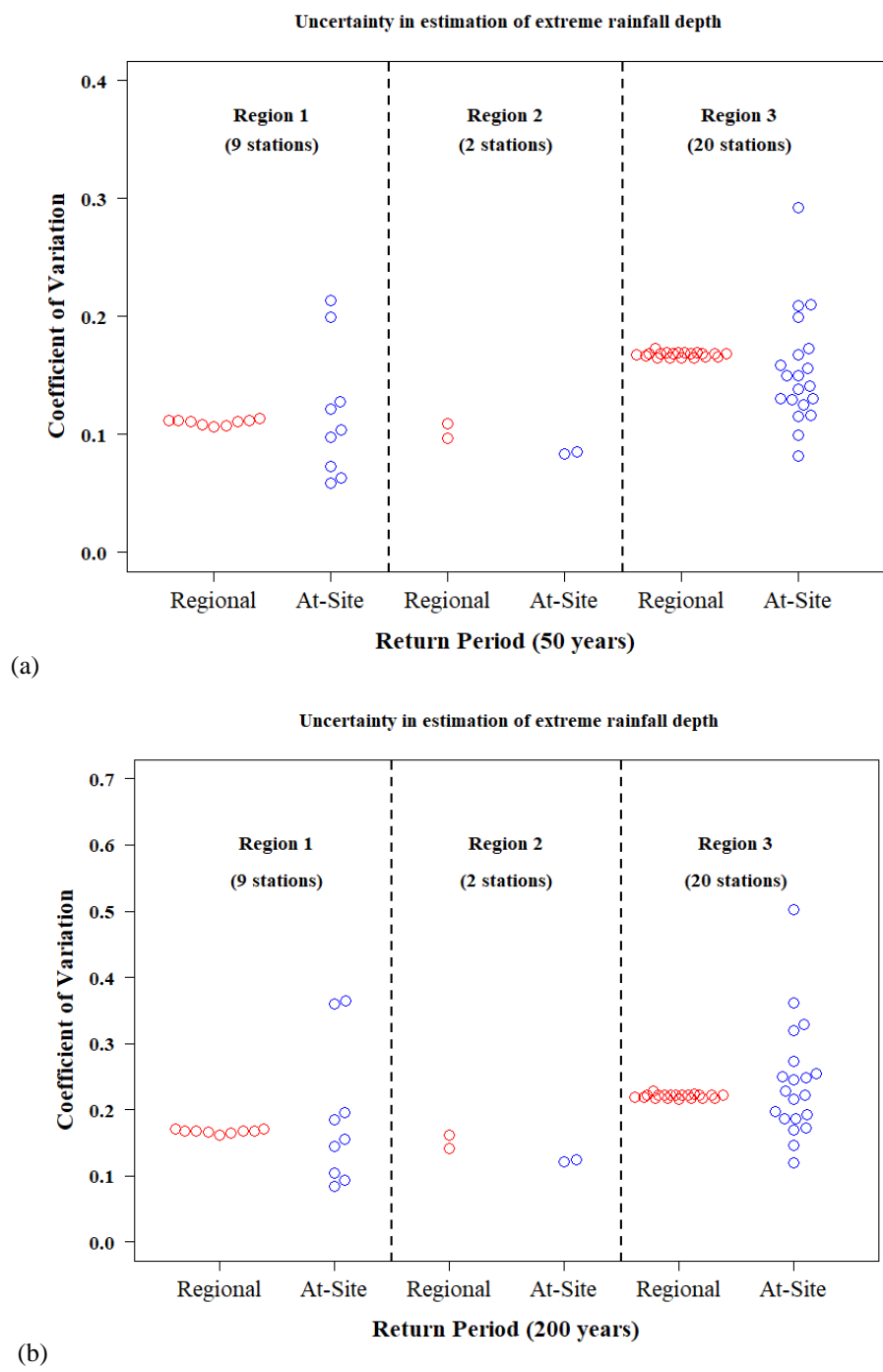
(a)

(b)

Figure 4. Uncertainty of extreme rainfall estimations in homogenous rainfall regions for (a) 50 and (b) 200 years return period

4.3. Confidence Interval based Uncertainty Analysis

To test the confidence intervals for each region, two sites were considered – the lowest and the highest discordant stations in each region – to see the effect on these two extremities. The lowest and highest discordant stations in region I are Dibrugarh and North Lakhimpur; in region II they are Cherrapunjee and Mawsynram, while in region III they are Silchar and Kailasahar. Region II consists of only two stations, and the assessment is made only between them. The Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) goodness-of-fit tests are used to examine how well the Empirical Cumulative Distribution Function (ECDF) and theoretical CDF fit the observed data. From the results of the goodness-of-fit test in Table 3, estimates for at-site and regional analysis for each region are explored, and the p-value results indicate that the observed data seems to come from a population with a PE3 distribution for region II and a GEV distribution for regions I and II. The p-values obtained further suggest that the highest discordant sites in each region seem to have less fit in comparison to the lowest discordancy sites when fitted through either at-site or regional analysis. For region II, the relative difference is not clearly distinguishable, but the p-values indicate a good fit in both sites.

Table 3. Goodness of-fit tests of lowest and highest discordant stations

Homogenous Region	Site	Frequency analysis method	AD test statistic	P-value	KS test statistic	P-value
I	Dibrugarh	At Site	0.273	0.976	0.150	0.983
		Regional	0.332	0.937	0.150	0.983
	North Lakhimpur	At Site	0.373	0.902	0.150	0.983
		Regional	0.552	0.715	0.200	0.832
II	Cherrapunjee	At Site	0.146	1.000	0.100	1.000
		Regional	0.229	0.993	0.150	0.983
	Mawsynram	At Site	0.192	0.998	0.100	1.000
		Regional	0.213	0.996	0.150	0.983
III	Silchar	At Site	0.206	0.997	0.100	1.000
		Regional	0.194	0.998	0.100	1.000
	Kailasahar	At Site	0.187	0.999	0.100	1.000
		Regional	0.336	0.934	0.150	0.983

The confidence intervals evaluated to estimate the uncertainty for the lowest and highest discordant stations in each region is presented in Figures 5 to 7. For Dibrugarh station in region I, the empirically determined precipitation quantiles all fall within the 95% confidence interval (CI) bounds for both MCS and normality assumptions. Highest discordant station of region I i.e., North Lakhimpur in Figure 5 shows wider CI bounds for at-site analysis with the empirical quantiles lying on the lower CI bounds of normality assumption. Thus, the 95% CI bounds are found less narrow and the uncertainty associated with estimation of at-site quantiles are relatively higher. In this region both the regional growth curve and at-site growth curve are well within the MCS CI and normality-based confidence intervals. For Cherrapunjee station in Figure 6, the upper limit of MC simulation bounds exceeds than normality assumptions, thereby indicating uncertainty of estimates to deviate from normality assumption. But as the regional quantile growth curve for the station is found to be well bounded and close to lower limit of both the normality and MCS CI bounds, there is no over-estimation of rainfall estimates. Whereas, the regional and at-site CI bounds for Mawsynram is well within the normality assumption bounds and do not show any significant difference. Thus, the uncertainty in quantiles for Mawsynram is least in region II. For station Kailasahar of region III in Figure 7, the confidence interval width was almost similar in both regional and at-site estimations. For Silchar with lowest discordancy in region III, the upper limits of MCS CI from regional approach have higher values thus providing greater widths for higher return periods. Thus, the uncertainty is seen to be more for lowest discordant station in this region. Though the empirical quantiles up to 8-year return periods are seen to fall on the lower limit of both the CI bounds, the observation circles are seen to return to position well within the MCS and normality CI bounds afterwards.

The regional estimates remain close to the lower limit of the CI bounds of both MC simulation and normality bounds, thereby indicating no overestimation in the quantile growth curve. Figure 5 to 7 overall indicates that the confidence intervals for regional quantiles calculated from MC simulation are narrower and follow normality assumptions in all three homogenous regions. The uncertainty is explored and presented for the discordant extremes in the regions. In comparison to at-site analysis, the regional approach was found to have low uncertainty in most of the stations.

To have an assessment of overall performance including all stations in a region, the average relative width (ARW) of the confidence intervals has been analysed and presented in Table 4. The MC simulation procedure is similarly followed for other stations in the region and confidence intervals are computed at each return period. Comparison to at-site estimates and confidence intervals for corresponding return periods are also done. The results show that, the regional analysis of rainfall estimates for region I produced narrower confidence intervals than at-site analysis. For region II, the widths of CI's across all stations in the region were slightly larger than at-site analysis CI's. But the CI's obtained for region II were relatively the least deviating among the three regions, thereby indicating that the GEV distribution satisfactorily describes the rainfall distribution in the region. For region III, the confidence intervals were better than those of at-site only for higher return periods of 500 and 1000 years. This indicates that the regional PE3 distribution was relatively less appropriate to at-site GEV distribution for the region III in producing accurate rainfall estimates. The clustered region III constitutes twenty stations which may be considered large, and the analysis has overestimation of regional growth curve from true at-site growth curve for some stations and underestimation on some stations. This may be due to the widespread location of the stations in both Brahmaputra and Barak basin.
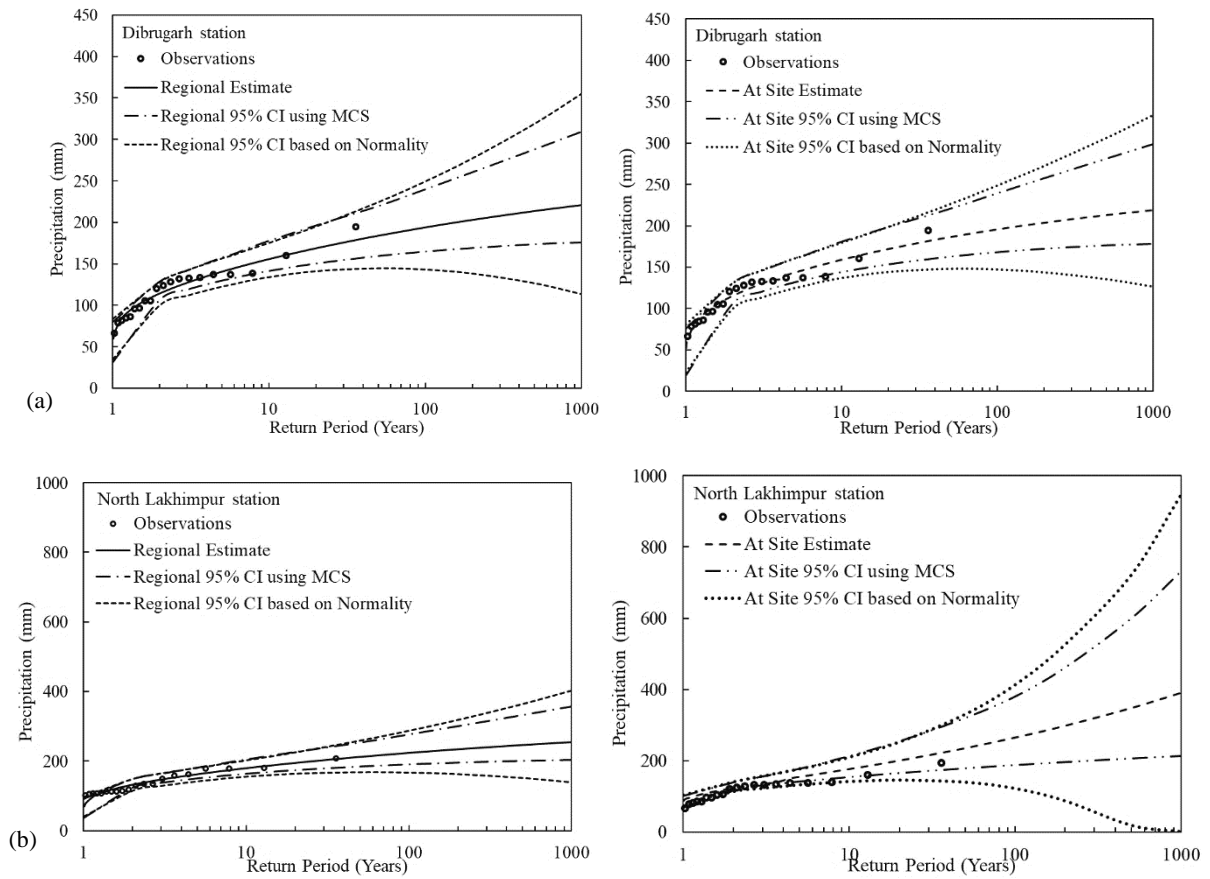
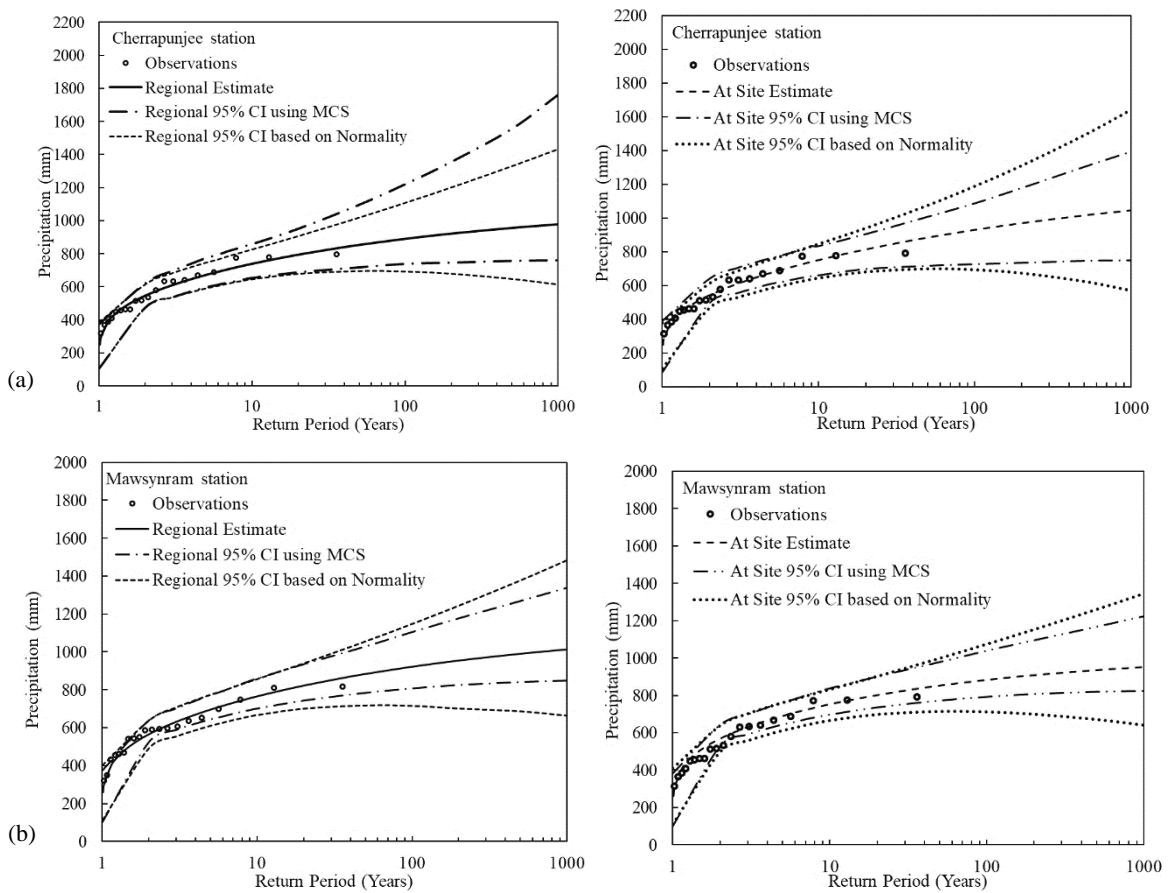**Figure 5. Comparison plot of confidence intervals for (a) lowest and (b) highest discordant stations for region I**

Figure 6. Comparison plot of confidence intervals for (a) lowest and (b) highest discordant stations for region II
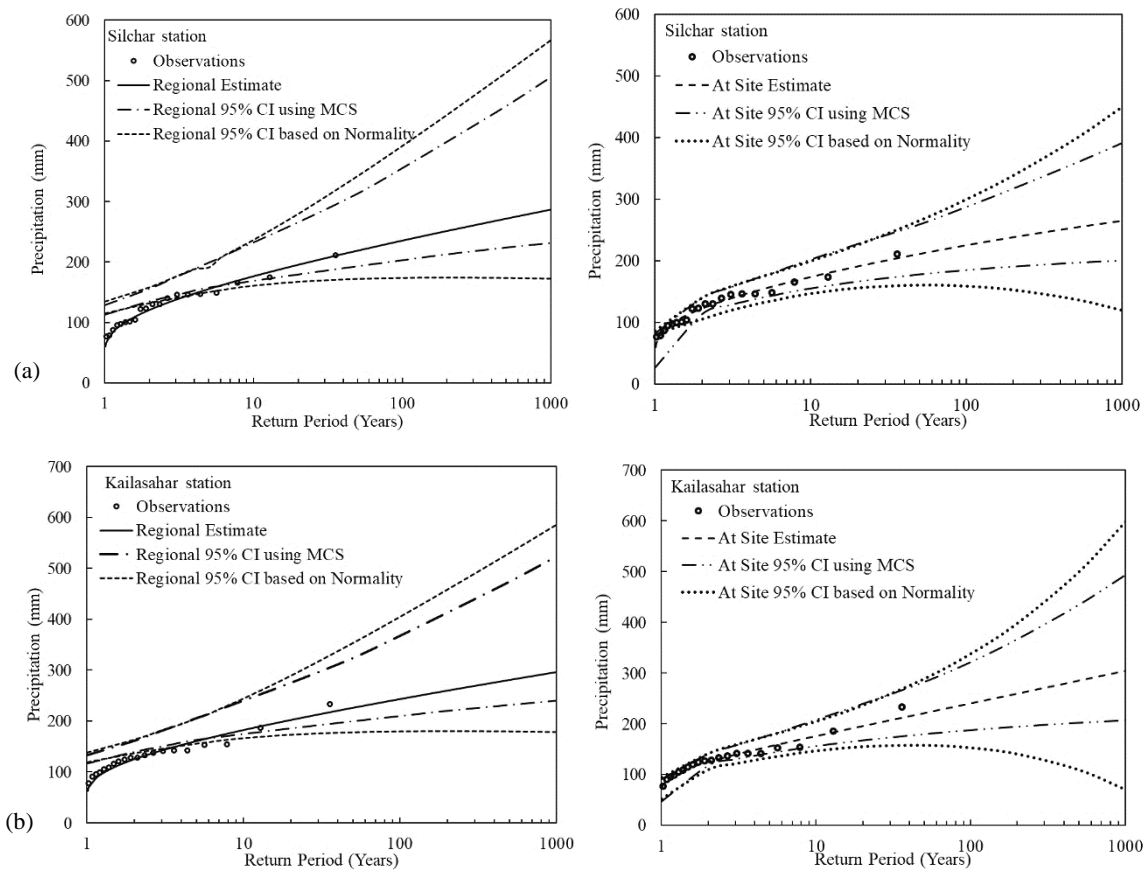
Figure 7. Comparison plot of confidence intervals for (a) lowest and (b) highest discordant stations for region III

Furthermore, altitude appears to have an impact on the quantile performance among the stations in the cluster regions. The region I, II and III has altitudinal difference between stations with highest and lowest altitude as 56, 88 and 1582 m. Thus, it can be seen that as the altitudinal variation in a cluster group increases, there is seen to observe a reduction in the efficiency of regional quantile estimates. Region III has the largest number of stations and constitutes the highest and lowest station altitudes in the study area, and hence rainfall estimates is found to vary in the region relatively more, leading to more uncertainty. Thus, the results indicate estimates from regional analysis is most accurate in region I, with slightly reduced performance in regions II and better performance for only higher return periods in region III. The uncertainty in regional analysis estimates is thus explored and with comparison to at-site approach in the delineated homogenous regions, is considered preferable.

Table 4. Average relative width (ARW) for regional and at-site rainfall quantiles

Region	No of stations	Frequency Analysis method	ARW for different return periods								
			2	5	10	20	50	100	200	500	1000
Region I	9	Regional	0.25	0.24	0.26	0.32	0.43	0.54	0.67	0.87	1.05
		At-Site	0.24	0.24	0.26	0.32	0.46	0.58	0.74	0.98	1.19
Region II	2	Regional	0.26	0.25	0.26	0.31	0.41	0.50	0.61	0.80	0.95
		At-Site	0.25	0.22	0.22	0.25	0.33	0.40	0.48	0.59	0.66
Region III	20	Regional	0.23	0.33	0.44	0.60	0.81	0.97	1.12	1.29	1.43
		At-Site	0.29	0.30	0.33	0.40	0.58	0.75	0.95	1.30	1.56

4.4. Uncertainty Analysis based on Information Transfer Index and MLE-GA

For regional analysis, the parameters of the selected distribution determined using L-Moments in each homogenous region is taken for lowest and highest discordant sites in each region. For uncertainty determination of the parameters, ungauged site (lowest and highest discordant sites of a homogenous region) are considered for assessment. Information transfer index (ITI) values were calculated based on equation number 9 to 11, and new random samples at all the sites

in the homogenous region are generated using parameters of the selected regional distribution. Then, random sample data of the original sample size as the considered site is generated based on ITI dependent weights using Equation 12. The parameters of the ITI based random sample is evaluated using L-Moments method and rainfall depths under different return periods (T= 10, 20, 50 and 200 years) is determined. The process is repeated 1000 times using Monte Carlo simulation approach, and quantiles under different are obtained. For each return period, estimated rainfall quantiles from 1000 Monte Carlo simulation are sorted and ranked and the 5th, 50th and 95th percentile is obtained. Similar procedure is applied for the other two types of weights viz. (i) distance-based dependent weights, and (ii) combination of ITI and distance-based weights to generate random sample from regional distribution at the considered site. Here, the distance is based on Euclidean distance and weights are determined using equation 13 and 14. In the present study, the stations of region 2 was not considered in the analysis, as ITI and distance-based weights for only two stations was not possible. Two new stations Karimganj and Lengpui was considered in the study for assessing the performance of the analyses. The sites were assigned to region III based on the Euclidean distance nearness of station attributes to the centroid of region III cluster group.

4.4.1. Regional Uncertainty

The ITI based weights gave better results in regional frequency analysis with estimates of rainfall performing better than at-site analysis for both least and highest discordant stations in region I; and for least discordancy of region III. Despite the fact that the fitted distribution for region I is both GEV distribution in both regional and at-site frequency analysis, L-Moment based regional rainfall estimates was found to clearly outperform the at-site estimates for all return periods and can be seen in Figure 8. This shows that, stations with least and highest discordancy in the region I gives better prediction with regional frequency analysis compared to at-site frequency analysis. The uncertainty in quantile estimates for regional analysis was observed the lowest with ITI based weighting and highest for distance-based weighting in all return period of the two regions except for highly discordant station in region III. This may be due to high regional absolute bias $A^R(F)$ of region III as presented in Table 2, as a higher value is suggestive of estimation of quantiles to be consistently high at some stations and low at others. The performance of the new method for two new stations i.e., Karimganj and Lengpui as in Figure 10, did not provide acceptable results as the regional estimates were significantly much higher compared to at-site estimates. One reason for this may be due to the data for the stations may not behave as the selected regional distribution for the homogenous group and may need to be included in clustering for proper allotment of homogenous group. But the performance of ITI based uncertainty compared to at-site was superior for Lengpui station, thereby suggesting the ITI based method of generating station data to be reliable and robust. Overall, for homogenous regions with low bias as in region I, the performance of ITI based uncertainty definitely outperformed at-site frequency analysis.

4.4.2. At-site Uncertainty

The ITI and distance-based weights for application of uncertainty in at-site frequency analysis was done considering 8 and 16 nearby surrounding stations. The grouping of stations into 8 and 16 stations was done by ranking and sorting the stations in terms of higher ITI value shared between the ungauged station (here least and highest discordant stations) for ITI dependent weights. For the distance-based and MLE-GA approach, the nearness to the study station of other stations was based on Euclidean distance. Figures 8 and 9 shows at-site estimates produced higher uncertainty in comparision to regional analysis estimates except for highest discordant station in region III. Applying the ITI based dependent weights, the uncertainty of at-site estimates was significantly reduced in comparisons to distance based at-site estimates. This result is suggestive of the fact that grouping of stations based on ITI yield much more reliable and correct information at the ungauged site. The uncertainty in rainfall estimates calculated by at-site estimates for all three approaches (ITI, distance-based and MLE-GA) is seen increasing with increase in return period for all stations, which is in agreement with the corresponding results of regional estimates. However, the rate of increase is reduced with inclusion of more extracted sites from 8 to 16. The uncertainty obtained for MLE-GA based estimates for both ITI and distance-based is found to have lower values compared to at-site ITI and distance-based estimates for return period of 50, 100 and 200 years for all regions. This suggests that the regional analysis comparision to at-sites estimates based on MLE method optimized by genetic algorithm estimates are more preferable. Though in many studies, it is generally observed that L-Moment method outperforms MLE method in regional frequency analysis, and MLE generally performs better with larger sample size. The present work found MLE to perform better than L-Moment method in at-site frequency analysis estimates and with low sample size of 20. For the two new sites Karimganj and Lengpui as presented in Figure 10, the MLE-GA method also performed better with least value of uncertainty for at-site estimates both for ITI and distance-based estimates. This suggests that the at-site estimates based on MLE-GA may serve as a better alternative for comparing regional frequency estimates.
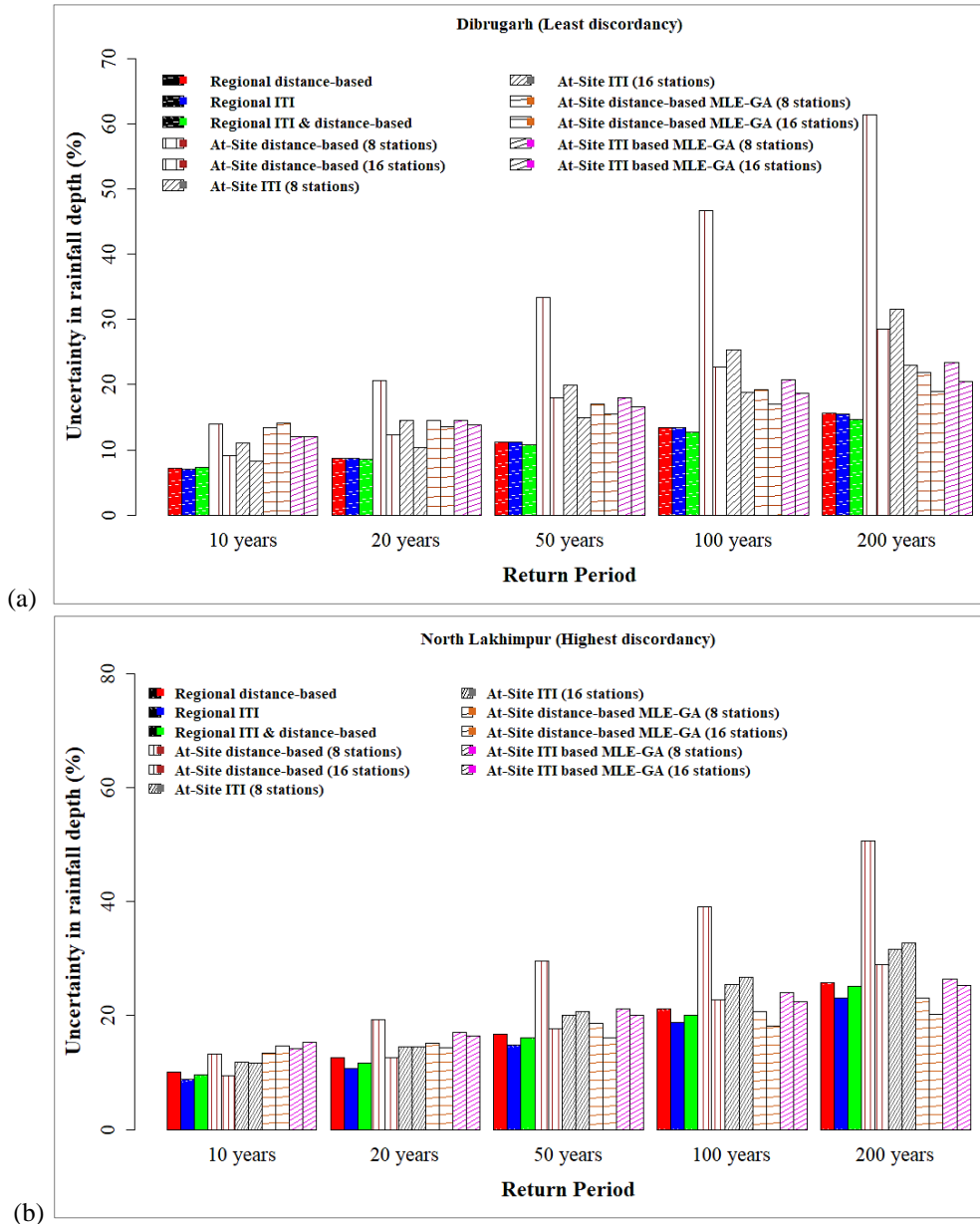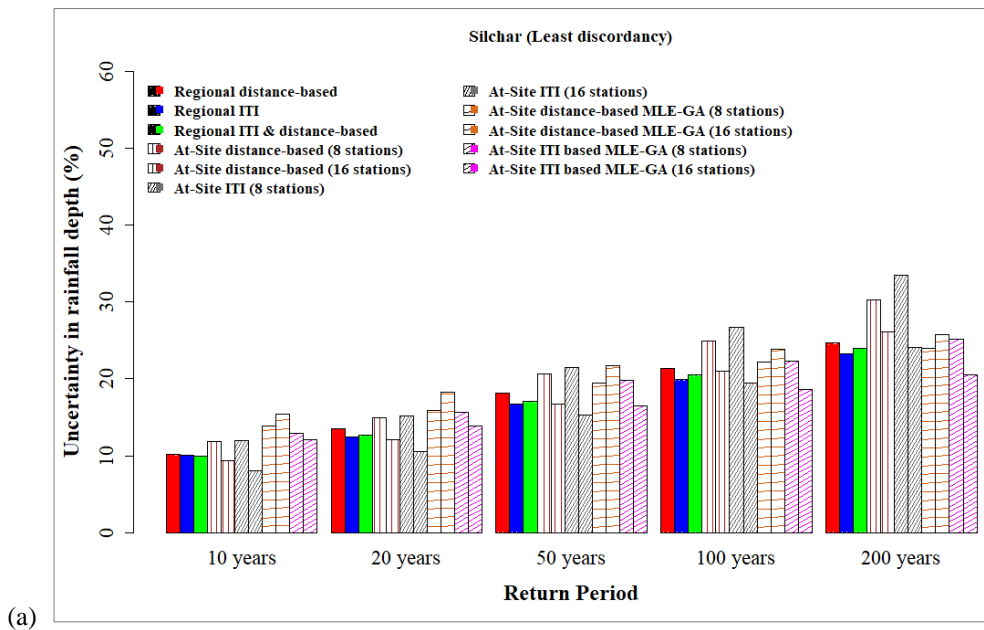
Figure 8. Uncertainty of extreme rainfall estimations for (a) least and (b) highest discordant stations in region I
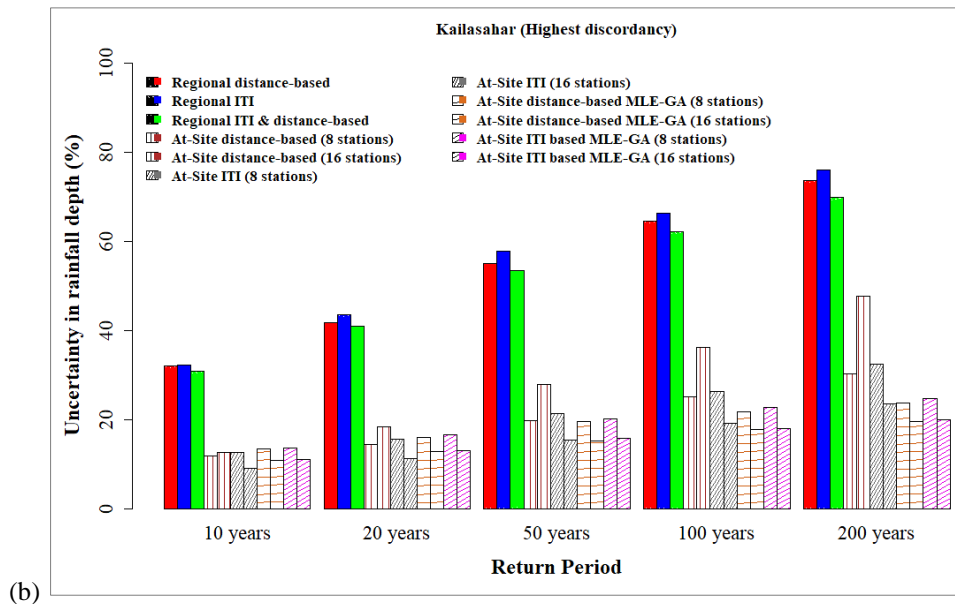
(b)

Figure 9. Uncertainty of extreme rainfall estimations for (a) least and (b) highest discordant stations in region III
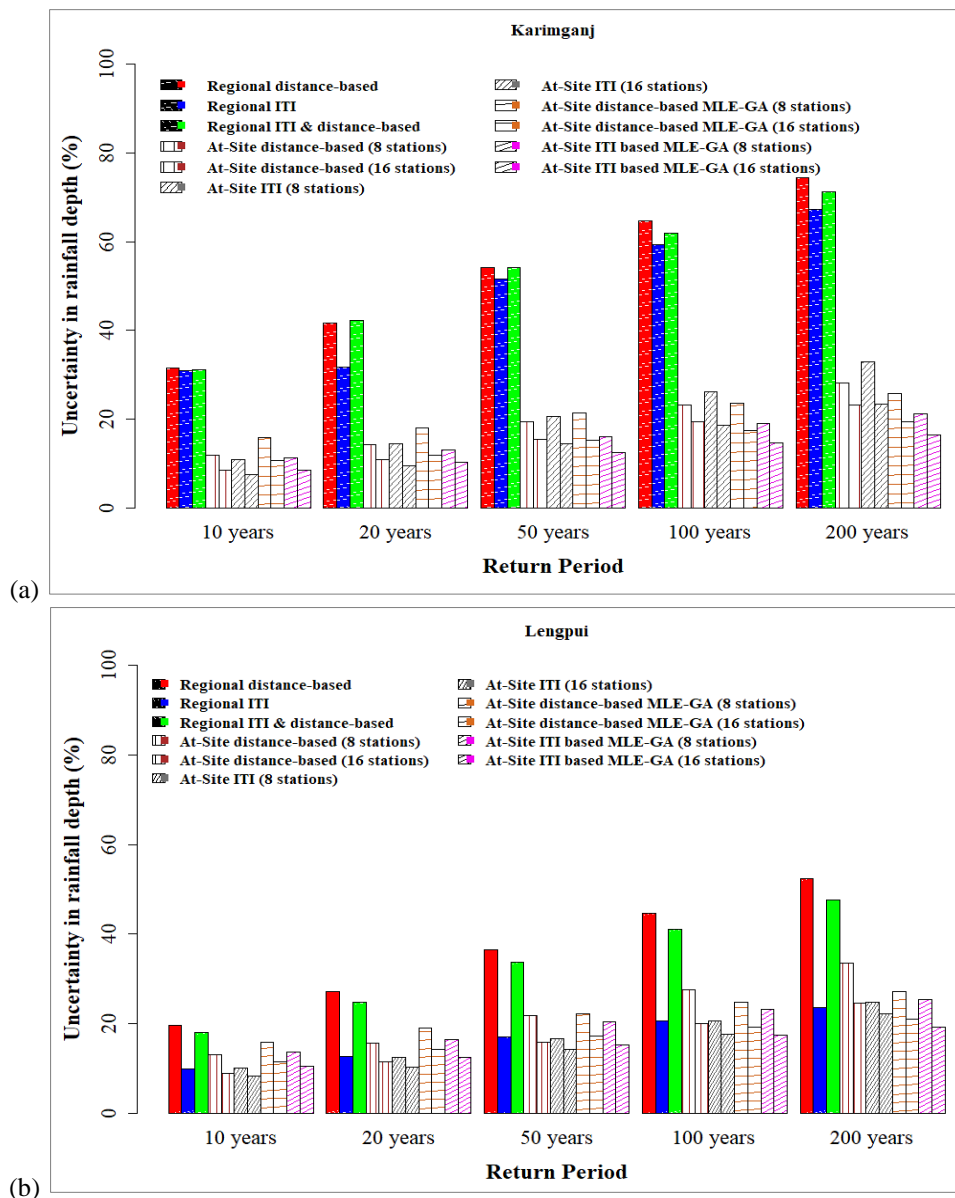

(a)

(b)

Figure 10. Uncertainty of extreme rainfall estimations for two ungauged stations (a) Karimganj (b) Lengpui not considered in grouping of homogenous regions

4.5. Comparison with Previously Done Similar Studies

Although this is the first study to present a comparison of regional and at-site rainfall estimates in the southern part of Brahmaputra and Barak region, some closely related studies in other parts of the world may be related for validation of the performance. For at-site analysis in Maryland, USA, Al Kazbaf and Bensi, 2021 [14] found that the choice of distribution and method of parameter estimation (LMOM, MLE and MOM) affected the shape and location of precipitation estimate hazard curve performance significantly. For regional analysis with GEV and GNO distributions, the effect of distribution choice had limited effects. This is in accordance with the results in the present study wherein GEV distribution of at-site analysis seems to perform better than regional analysis by PE3 distribution for stations in region III. While for region I where both at-site and regional analysis are based on GEV distribution the regional analysis performed better in uncertainty. So, the choice of distribution is an important parameter in regional frequency analysis. Also, the parameter estimation method MLE-GA was found to perform better for highly discordant site and lower to least discordant site in region III compared to regional analysis. Yin et al. 2016 [33] compared the accuracy of regional and at-site quantiles of Yangtze River delta region based on RMSE and obtained lower RMSE for regional analysis for longer return periods. Li et al. 2019 [15] considered the lowest and highest discordant stations in nine homogenous regions of Sichuan province, China and found that stations with lowest discordancy had smaller differences of design rainfall values for both regional and at-site frequency analysis compared to stations with highest discordancy in the region. This is seen in the study presented, with larger difference for highest discordancy sites in region I and III. Zhou et al. 2014 [12] compared the MLE and L-Moment method for annual extreme precipitation estimates in Taihu basin of China for GEV and PE3 distribution and found MLE to provide unreasonable higher estimates compared to L-Moment estimates. In the present study, MLE method optimised using GA gave reasonable estimates for at-site analysis using GEV distributions and performed better to at-site estimates based on L-Moments method. But the precipitation estimates based on L-Moment regional frequency analysis performed superior to at-site analysis for both L-Moment and MLE-GA estimation methods for most stations. The MLE-GA estimates for at-site analysis for the stations in the homogenous regions from both ITI and distance-based estimates was accurately estimated with observance of no unreasonable result.

5. Summary and Conclusions

The study focused on the performance of extreme rainfall quantiles for homogenous regions delineated by genetic algorithm-based clustering. Uncertainty and accuracy assessment was investigated for the selected frequency distributions of the derived homogenous regions. Two distributions GEV and PE3 were found to satisfactorily define the annual extreme rainfall behavior in the study area. Regional growth curves of GEV and PE3 distributions from regional frequency analysis gave minimal bias and least deviation in all three regions. The uncertainty associated with regional rainfall quantiles is then reported using coefficient of variation C_v, and is found to be consistent and fairly low for all considered stations in the regions. Whereas analysis from at-site quantiles for the stations were seen to be highly inconsistent and produced higher values of C_v with increase in return periods. Results obtained suggest consistency in uncertainty of rainfall estimates for regional analysis, with larger variation in at-site analysis.

Results of uncertainty for regional quantiles of lowest and highest discordant stations in all three delineated homogenous rainfall regions did not seem to differ distinctly, and were within the confidence limits of both Monte Carlo simulation and normality assumptions. Whereas the uncertainty of quantiles estimated from at-site analysis increased after return period of 100 years in regions I and III. Results also show the uncertainty associated with rainfall quantiles derived from Monte Carlo simulation to follow normal distribution, and hence the regional rainfall quantiles were satisfactorily accurate. Region III comprised of 20 stations and were widely spread across both Brahmaputra and Barak basin. Growth curve in this region gave higher absolute relative bias for return periods up to 200 years, and may be attributed to a large number of stations with distinctly varying altitudes. Further, altitude seems to have influence on regional frequency analysis in the northeast region. As the altitudinal variation of stations for a cluster group increased, reduction in the accuracy of estimated regional quantile estimates was observed.

An assessment of overall performance of a homogenous region with average relative width of confidence interval showed that regional analysis produced narrower confidence intervals than at-site analysis in region I. While the average relative width (ARW) of region II and III were not as good as region I and had slightly higher values for regional analysis. But, the regional estimates was found to be better at higher return periods of 500 and 100 years in region III. Based on the ARW results, genetic algorithm based clustering approach is found to be a robust method in determining homogeneous regions and hence can be applied in determining reliable and accurate rainfall estimates for any study region.

The ITI-based weights produced superior results in regional frequency analysis, with rainfall estimates outperforming on-site analysis for both the least and most discordant stations in region I, as well as the least discordant stations in region III. Except for the most discordant station in area III, the uncertainty was determined to be lowest when ITI weighting was applied and highest when distance weighting was utilized. Regional analysis outperformed at-

site frequency analysis based on ITI-dependent derived weights for homogeneous regions with little bias. For all regions, the uncertainty in quantile estimates using at-site analysis was consistently greater than the uncertainty in quantile estimates using a distance-based weighting technique. This finding suggests that ITI provides more valuable information across sites and can combine sites to provide more accurate information at an ungauged site. Additionally, the uncertainty associated with MLE-GA-based at-site estimates of ITI and distance-based estimates is shown to be lower and more preferred than that associated with at-site L-Moments-based ITI and distance-based estimates. While the MLE-GA approach performed best for the two new unmeasured sites Karimganj and Lengpui, it did so with the least uncertainty, indicating that at-site estimates based on the MLE-GA method may be a preferable choice for comparison. The results of this study will be helpful in promoting the differences between regional and at-site frequency estimates in the context of hydrological frequency analysis. At the same time, it will be helpful in assisting decisions related to risk and hazard mitigation of extreme rainfall events in the northeast region of India.

6. Declarations

6.1. Author Contributions

N.D. contributed to the conception, design and write-up of the manuscript; P.R. and P.C. guided and supervised the research work; S.A. reviewed and edited the first draft of the manuscript. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Acknowledgements

We are grateful to Regional Meteorological Centre Guwahati, India for supply of data for the study region.

6.5. Conflicts of Interest

The authors declare no conflict of interest.

7. References

[1] Svensson, C., and D. A. Jones. "Review of Rainfall Frequency Estimation Methods." Journal of Flood Risk Management 3, no. 4 (2010): 296–313. doi:10.1111/j.1753-318X.2010.01079.x.

[2] Tasker, Gary, J. R. M. Hosking, and J. R. Wallis. Regional Frequency Analysis: An Approach Based on L-Moments. Journal of the American Statistical Association. Vol. 93. University Press, Cambridge, 1998. doi:10.2307/2669866.

[3] Hazarika, Jayshree, and Arup Kumar Sarma. "Importance of Regional Rainfall Data in Homogeneous Clustering of Data-Sparse Areas: A Study in the Upper Brahmaputra Valley Region." Theoretical and Applied Climatology 145, no. 3–4 (2021): 1161–75. doi:10.1007/s00704-021-03686-x.

[4] Debbarma, Nilotpal, Parthasarathi Choudhury, and Parthajit Roy. "Identification of Homogeneous Rainfall Regions Using a Genetic Algorithm Involving Multi-Criteria Decision Making Techniques." Water Science and Technology: Water Supply 19, no. 5 (2019): 1491–99. doi:10.2166/ws.2019.018.

[5] Goyal, Manish Kumar, and Vivek Gupta. "Identification of Homogeneous Rainfall Regimes in Northeast Region of India Using Fuzzy Cluster Analysis." Water Resources Management 28, no. 13 (2014): 4491–4511. doi:10.1007/s11269-014-0699-7.

[6] Deka, S, M Borah, and S C Kakaty. "Distributions of Annual Maximum Rainfall Series of North-East India." European Water 1, no. 27/28 (2009): 3–14.

[7] Deka, Surobhi, Munindra Borah, and Sarat Chandra Kakaty. "Statistical Analysis of Annual Maximum Rainfall in North-East India: An Application of LH-Moments." Theoretical and Applied Climatology 104, no. 1–2 (2011): 111–22. doi:10.1007/s00704-010-0330-7.

[8] Hailegeorgis, Teklu T., and Knut Alfredsen. "Regional Flood Frequency Analysis and Prediction in Ungauged Basins Including Estimation of Major Uncertainties for Mid-Norway." Journal of Hydrology: Regional Studies 9 (2017): 104–26. doi:10.1016/j.ejrh.2016.11.004.

[9] Souza, Rezende de, Merwade G., Coutinho de Oliveira V., and Viola L. F. "M." R., and Sa Farias, M. ``Regional Flood Frequency Analysis and Uncertainties: Maximum Streamflow Estimates in Ungauged Basins in the Region of Lavras, MG, Brazil 197 (2021): 10497. https://doi.org/10.1016/j.catena.2020.104970.

[10] Liang, Yuyin, Shuguang Liu, Yiping Guo, and Hong Hua. "L-Moment-Based Regional Frequency Analysis of Annual Extreme Precipitation and Its Uncertainty Analysis." Water Resources Management 31, no. 12 (2017): 3899–3919. doi:10.1007/s11269-017-1715-5.

[11] Zhu, Qian, Haiting Gu, Xujie Zhang, and Yue-Ping Xu. "Parameter Uncertainty and Nonstationarity in Regional Rainfall Frequency in Qu River Basin, East China." Journal of Hydrologic Engineering 21, no. 5 (2016): 04016008. doi:10.1061/(asce)he.1943-5584.0001355.

[12] Zhou, Zhengzheng, Shuguang Liu, Hong Hua, Chang-Shian Chen, Guihui Zhong, Hejuan Lin, and Chien-Wei Huang. "Frequency Analysis for Predicting Extreme Precipitation in Changxing Station of Taihu Basin, China." Journal of Coastal Research 68 (2014): 144–51. doi:10.2112/si68-019.1.

[13] Merz, Bruno, and Annegret H. Thieken. "Separating Natural and Epistemic Uncertainty in Flood Frequency Analysis." Journal of Hydrology 309, no. 1–4 (2005): 114–32. doi:10.1016/j.jhydrol.2004.11.015.

[14] Kajbaf, Azin Al, and Michelle B. "Assessment of Uncertainty in Regional and At-Site Precipitation Frequency Analysis for the Localized Region of Ellicott City, Maryland." Natural Hazards 108, no. 3 (2021): 2513–41. doi:10.1007/s11069-021-04786-7.

[15] Li, Mengrui, Xiaodong Li, and Tianqi Ao. "Comparative Study of Regional Frequency Analysis and Traditional At-Site Hydrological Frequency Analysis." Water (Switzerland) 11, no. 3 (2019): 486. doi:10.3390/w11030486.

[16] Goudenhoofdt, Edouard, Laurent Delobbe, and Patrick Willems. "At-Site and Regional Frequency Analysis of Extreme Precipitation from Radar-Based Estimates." Hydrology and Earth System Sciences, 2017, 1–26. doi:10.5194/hess-2017-150.

[17] Cullen, Alison C., and H. Christopher Frey. Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs, Springer, Berlin, (1999).

[18] Zhang, Wei, and Tian Li. "The Influence of Objective Function and Acceptability Threshold on Uncertainty Assessment of an Urban Drainage Hydraulic Model with Generalized Likelihood Uncertainty Estimation Methodology." Water Resources Management 29, no. 6 (2015): 2059–72. doi:10.1007/s11269-015-0928-8.

[19] Sha, Jian, Zeli Li, Dennis P. Swaney, Bongghi Hong, Wei Wang, and Yuqiu Wang. "Application of a Bayesian Watershed Model Linking Multivariate Statistical Analysis to Support Watershed-Scale Nitrogen Management in China." Water Resources Management 28, no. 11 (2014): 3681–95. doi:10.1007/s11269-014-0696-x.

[20] Zheng, Yi, and Feng Han. "Markov Chain Monte Carlo (MCMC) Uncertainty Analysis for Watershed Water Quality Modeling and Management." Stochastic Environmental Research and Risk Assessment 30, no. 1 (2016): 293–308. doi:10.1007/s00477-015-1091-8.

[21] Nguyen, Duc Hai, Seon Ho Kim, Hyun Han Kwon, and Deg Hyo Bae. "Uncertainty Quantification of Water Level Predictions from Radar‑based Areal Rainfall Using an Adaptive MCMC Algorithm." Water Resources Management 35, no. 7 (2021): 2197‑2213. doi:10.1007/s11269-021-02835-1.

[22] Ridolfi, E., M. Rianna, G. Trani, L. Alfonso, G. Di Baldassarre, F. Napolitano, and F. Russo. "A New Methodology to Define Homogeneous Regions through an Entropy Based Clustering Method." Advances in Water Resources 96 (2016): 237–50. doi:10.1016/j.advwatres.2016.07.007.

[23] Harmancioglu, Nilgun, and Vujica Yevjevich. "Transfer of Hydrologic Information among River Points." Journal of Hydrology 91, no. 1–2 (1987): 103–18. doi:10.1016/0022-1694(87)90131-4.

[24] Rajsekhar, Deepthi, Ashok K. Mishra, and Vijay P. Singh. "Regionalization of Drought Characteristics Using an Entropy Approach." Journal of Hydrologic Engineering 18, no. 7 (2013): 870–87. doi:10.1061/(asce)he.1943-5584.0000683.

[25] Yang, Yujuin, and Donald H. Burn. "An Entropy Approach to Data Collection Network Design." Journal of Hydrology 157, no. 1–4 (1994): 307–24. doi:10.1016/0022-1694(94)90111-2.

[26] Du, Hong, Jun Xia, and Sidong Zeng. "Regional Frequency Analysis of Extreme Precipitation and Its Spatio-Temporal Characteristics in the Huai River Basin, China." Natural Hazards 70, no. 1 (2014): 195–215. doi:10.1007/s11069-013-0808-6.

[27] Jin, Xiaoli, Chong Yu Xu, Qi Zhang, and V. P. Singh. "Parameter and Modeling Uncertainty Simulated by GLUE and a Formal Bayesian Method for a Conceptual Hydrological Model." Journal of Hydrology 383, no. 3–4 (2010): 147–55. doi:10.1016/j.jhydrol.2009.12.028.

[28] Uboldi, F., A. N. Sulis, C. Lussana, M. Cislaghi, and M. Russo. "A Spatial Bootstrap Technique for Parameter Estimation of Rainfall Annual Maxima Distribution." Hydrology and Earth System Sciences 18, no. 3 (2014): 981–95. doi:10.5194/hess-18-981-2014.

[29] Lall, Upmanu, and Ashish Sharma. "A Nearest Neighbor Bootstrap for Resampling Hydrologic Time Series." Water Resources Research 32, no. 3 (1996): 679–93. doi:10.1029/95WR02966.

[30] Xu, Sudong, and Wenrui Huang. "Estimating Extreme Water Levels with Long-Term Data by GEV Distribution at Wusong Station near Shanghai City in Yangtze Estuary." Ocean Engineering 38, no. 2–3 (2011): 468–78. doi:10.1016/j.oceaneng.2010.11.022.

[31] Ljung, G. M., and G. E.P. Box. "On a Measure of Lack of Fit in Time Series Models." Biometrika 65, no. 2 (1978): 297–303. doi:10.1093/biomet/65.2.297.

[32] Atiem, Isameldin A., and Nilgün B. Harmancioğlu. "Assessment of Regional Floods Using L-Moments Approach: The Case of the River Nile." Water Resources Management 20, no. 5 (2006): 723–47. doi:10.1007/s11269-005-9004-0.

[33] Yin, Yixing, Haishan Chen, Chong Yu Xu, Wucheng Xu, Changchun Chen, and Shanlei Sun. "Spatio-Temporal Characteristics of the Extreme Precipitation by L-Moment-Based Index-Flood Method in the Yangtze River Delta Region, China." Theoretical and Applied Climatology 124, no. 3–4 (2016): 1005–22. doi:10.1007/s00704-015-1478-y.