CHARACTERIZATION OF DAILY PRECIPITATION EXTREMES OVER INDIA

DOCTORAL THESIS

By

NEHA GUPTA

(2017CEZ0006)

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY

ROPAR, PUNJAB 140001, INDIA

DECEMBER, 2023

CHARACTERIZATION OF DAILY PRECIPITATION EXTREMES OVER INDIA

A Thesis Submitted In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

by

Neha Gupta

(2017CEZ0006)

Department of Civil Engineering

Indian Institute of Technology, Ropar

December 2023

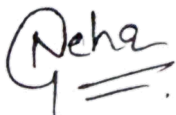Neha Gupta: *Characterization of daily precipitation extremes over India*

Dedicated

To

Family & Friends

DECLARATION OF ORIGINALITY

I declare that this work entitled "Characterization of daily precipitation extremes over India" has not previously been accepted in substance for any degree and is not being simultaneously submitted in candidature for any other degree. This thesis is being submitted in partial fulfillment of the requirements for the degree of Ph.D. in Civil Engineering. This thesis is the result of my own independent investigation, except where otherwise stated. I have acknowledged all the other sources by stating the references explicitly. I declare that any idea/data/fact/source stated in my thesis has not been fabricated/falsified/misrepresented. All the principles of academic honesty and integrity have been followed. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited or from whom proper permission has not been taken when needed. I hereby give consent for my thesis, if accepted, to be available online in the Institute's Open Access repository and for inter-library loan, and for the title and abstract to be made available to outside organizations.

Signature

Name: Neha Gupta

Entry Number: 2017CEZ0006

Program: Ph.D.

Department: Civil Engineering

Indian Institute of Technology

Ropar, Punjab, 140001

Date: December, 2023

<u>CERTIFICATE</u>

This is to certify that the thesis entitled **Characterization of daily precipitation extremes over India**, submitted by **Neha Gupta** for the award of the degree of **Doctor of Philosophy** of Indian Institute of Technology Ropar, is a record of bonafide research work carried out under my guidance and supervision. To the best of my knowledge and belief, the work presented in this thesis is original and has not been submitted, either in part or full, for the award of any other degree, diploma, fellowship, associateship, or similar title of any university or institution.

In my opinion, the thesis has reached the standard fulfilling the requirements of the regulations relating to the Degree.

Signature of the Supervisor
Name: Dr. Sagar Rohidas Chavan
Department: Civil Engineering
Indian Institute of Technology
Ropar, Punjab
Date: December, 2023

ACKNOWLEDGEMENT

LAY SUMMARY

The significance of information regarding the frequency and magnitude of extreme precipitation is paramount in hydroclimatic studies. Understanding the magnitude and frequency of precipitation extremes is of broad interest to both scientific and managerial communities due to its impact on the design of major civil engineering infrastructure, human and aquatic habitats, and water resource management. This critical information can be acquired by analyzing the behavior and nature of the upper tail of probability distributions that are suitable to represent the daily precipitation datasets.

The selection of the most appropriate probability distribution is of utmost importance, as an incorrect choice can result in either overestimating or underestimating the design precipitation for hydraulic infrastructure, leading to adverse consequences. Heavy-tail distributions have proven to be more suitable for representing precipitation extremes as opposed to distributions with exponential tails. A heavy-tail distribution indicates a higher likelihood of extreme events compared to distributions like Exponential, Gamma, and Gumbel. Analyzing tail behavior helps in assessing the likelihood of extreme events.

Given the limitations of simple parametric models in equally fitting the entire range of data, hydrologists, climatologists, and statisticians have developed advanced approaches for distribution fitting. These approaches prioritize the tails of distributions over the entire dataset, contributing to a better understanding of the tail behavior within hydroclimatological datasets.

However, quantifying and assessing upper tail behavior, such as tail heaviness, is a complex task. Typically, three main methods are employed for this purpose:

1. Threshold-based approaches (e.g., Block Maxima, Peak-over-Threshold)

2. Quantitative or scalar diagnostic indicators (e.g., shape parameters of common distributions like GEV, GP, obesity index (OB), Gini index, Upper tail ratio (UTR), and surprise factor)

3. Graphical approach (e.g., Mean excess plots, log-log plots, and Hill ratio plots).

The research presented in this thesis explores various aspect of the approaches mentioned and ask the following overarching question: How valid is the choice of the threshold in threshold-based approaches? Can scalar indicators help overcome the limitations of threshold-based approaches? Should we rely solely on a single graphical tool to diagnose tail behavior, or is a combination of tools preferable? What is the actual risk associated with varying tail behavior, and what is the most effective method for assessing tail risk information?

The answers to these questions are explored in the context of Indian precipitation extremes.

ABSTRACT

Daily precipitation extremes play a critical role in the hydrological planning and design of major water control structures and are expected to show a changing tendency over time due to climate change. The magnitude and frequency of extreme precipitation can be assessed by studying the upper tail behavior of probability distributions of daily precipitation datasets. These unexpected low-probability events lie within the tail part and have unprecedented consequences, underscoring the importance of their accurate estimation and prediction. The primary challenge with conventional distribution fitting approaches arises from the limited availability of data pertaining to extreme events. Because of this constraint, these methods struggle to effectively model the tails of daily precipitation data, often categorizing extreme precipitation events as improbable outliers. Consequently, this leads to an underestimation of their likelihood of occurrence. An appealing approach to overcome this impediment is the assessment of the tail behavior using some modern-day techniques like advanced threshold-based approaches, Quantitative or Scalar diagnostic tools, and Graphical approaches. The thesis concerns the development of novel approaches that can assess the tail behavior of precipitation extremes, thereby overcoming the limitations associated with old approaches. Characterizing the tail behavior of the daily precipitation finds use in the design and risk assessment of water control structures, economic evaluation of flood protection projects, flood insurance assessment, land use planning and management, and operation of irrigation projects. In the context of climate change, a better understanding of the climate extremes in terms of their frequency, magnitude, and spatial and temporal variation is necessary to evaluate the implications for risk and resilience. Hence, this thesis presents the study carried out to deliver a comprehensive assessment of extreme climatic conditions in India using some novel advanced approaches.

The initial part of the thesis is devoted to the application of threshold-based approaches to characterize the daily precipitation datasets over India. The investigation is carried out using the approach proposed by Papalexiou et al. (2013), where a Probability ratio mean square error (PRMSE) norm, is used to identify the best-fitting distribution to the tails of daily precipitation. Analysis related to the spatial-temporal change in the tail behavior of daily precipitation over India from pre- to post-1970 time periods as per the global climatic shift is done. The results indicate that the heavy-tailed distribution fits the tails of daily precipitation for the majority of the grids over India and an increase in the heaviness of tails of daily precipitation data over India from pre- to post-1970 time periods is observed.

In the second part, an empirical index known as the "Obesity index" (OB) that can provide a quantitative comparison between two distributions by alleviating the shortcomings associated with the threshold-based approaches is developed. The OB-based approach is applied to discern the probability distribution of daily gridded precipitation data for historical (1951–2004) and future (2006–2099)

periods over India into light- and heavy-tailed. Future projections of daily precipitation were obtained by downscaling simulations of the Coordinated Regional Climate Downscaling Experiment. Subsequently, a comparative analysis between the OB-based approach and threshold-based approaches by Nerantzaki and Papalexiou and Papalexiou et al. was conducted. Finally, the application of the OB-based approach is extended to characterize daily precipitation in Indian Meteorological subdivisions. Furthermore, we explored the dependence of the OB on the elevation of grids. Results indicated the applicability of heavy-tailed distributions in the representation of daily precipitation over India and suggested an OB-based approach as a good alternative diagnostic tool for assessing tail behavior.

The development of the Comprehensive Decision support system (DSS) was uptaken in the next part of the work, where several advanced graphical methods like Concentration profile (CP) plot, Concentration adjusted expected shortfall (CAES) plot, Zenga plot, Maximum-to-Sum plot, and Discriminant Moment ratio plot were incorporated together. Incorporation of advanced tools alleviates the limitations like lack of efficient segregation of the Lognormal distribution from the Regularly varying and Subexponential distribution families, associated with the conventional DSS. The robustness of the proposed DSS is established through a simulation experiment while the application was done to characterize the tails of daily gridded precipitation data over India. It is observed that about 98% of grids over India exhibit distributions from heavy-tailed families, which is of paramount concern as this shows higher frequency and magnitude of extreme over the Country.

The final portion of the study is aimed at discussing a comprehensive framework for estimating the risk associated with the tails of the daily precipitation datasets. Inferences from the novel approach like Concentration Profile (CP) are combined with the standard results from utility theory to develop a tool known as a Concentration Map (CM), that assesses the riskiness of datasets taking into account the variability of the larger and most relevant events. Risk embedded into the tails was evaluated for gridded precipitation datasets for the historical time period (1901–2019) from Indian Meteorological Department (IMD), while the simulations from 16 General Circulation Models (GCMs) participating in the Coupled Model Intercomparison Project phase 6 (CMIP6) under four Shared Socioeconomic Pathway (SSPs), namely, SSP126, SSP245, SSP370 and SSP585 are considered for future (2020-2100). The potential spatial and temporal variation of tail risk is done by comparing tail risk estimates from CMIP6 experiments (SSP126, SSP245, SSP370, SSP585) with historical datasets. Results highlight an overall increase in tail risk, particularly in scenarios indicative of anthropogenic influences, Furthermore, the analysis is extended to assess the variation in the embedded tail risk associated with daily precipitation datasets across different meteorological subdivisions and climate zones based on a Köppen-Geiger (KG) climate classification system, during different periods.

In a changing climate, understanding extreme precipitation events and their associated risks has become increasingly crucial. This study has employed advanced techniques and tools to illuminate the complexities of India's climate. The findings of this research can serve as a valuable guide for

policymakers in preparing for a future marked by more frequent and severe weather events. Local decision-makers can use the information provided in this thesis to effectively address the challenges presented by shifting climate patterns and formulate appropriate adaptation strategies in their respective regions.

LIST OF PUBLICATIONS

Journal papers

Gupta, N., & Chavan, S. R. (2024). Assessing future changes in daily precipitation tails over India: insights from multimodel assessment of CMIP6 GCMs. *Theoretical and Applied Climatology*, 1-19**.** *https://doi.org/10.1007/s00704-024-04849-2*

Gupta, N. & Chavan, S. R. (2023). A comprehensive decision support system for characterization of tails of probability distributions of daily precipitation. *Journal of Hydrology*. *https://doi.org/10.1016/j.jhydrol.2023.130282*

Gupta, N. & Chavan, S. R. (2023). Investigating the tail behavior and associated risk with daily discharges in South Indian Rivers. *Stochastic Environmental Research and Risk Assessment*, *37*, 3383–3399. *https://doi.org/10.1007/s00477-023-02453-w*

Gupta, N. & Chavan, S. R. (2022). Characterizing the tail behavior of daily precipitation probability distributions over India using the obesity index. *International Journal of Climatology*, *42*(4), 2543-2565. *https://doi.org/10.1002/joc.7380*

Chavan, S. R. & **Gupta, N**. (2022). On Investigation of Magnitude and Frequency of Annual Maximum Daily Precipitation in the Catchment of Bhakra Dam. INCOLD Journal (A Half Yearly Technical Journal of Indian Committee on Large Dams), 11(2), 18-28. Online ISSN: 2278-4691.

Gupta, N. & Chavan, S. R. (2021). Assessment of temporal change in the tails of the probability distribution of daily precipitation over India due to climatic shift in the 1970s. *Journal of Water and Climate Change*, *12*(6), 2753-2773. *https://doi.org/10.2166/wcc.2021.008*.

Gupta, N., & Chavan, S. R. (2023). Assessment of changes in monthly streamflow using innovative polygon trend analysis in the South Indian Rivers. *Arabian Journal of Geosciences*, *16*(12), 657.

Gupta, N. & Chavan, S. R. Assessment of embedded risk in precipitation tails over India through Concentration Profiles: A multi-model assessment from CMIP6 experiments (Under Review in Stochastic Environmental Research and Risk Assessment).

International and National Conferences

Gupta, N. & Chavan, S. R. (2023, July). Assessment of the Embedded Risk Associated with the Daily Precipitation Dataset over India, AOGS 2023, Singapore, 30 July – 4 Aug 2023, HS28-A016.

Gupta, N. & Chavan, S. R. (2023, April). Assessing daily precipitation tails over India under changing climate, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-13328, https://doi.org/10.5194/egusphere-egu23-13328

Gupta, N. & Chavan, S. R. (2022, December). Graphical trend analysis of streamflow in Cauvery River basin, India, International Conference on Hydraulics, Water Resources and Coastal Engineering, PEC Chandigarh, India. December 22-24, 2022, *ISH - HYDRO 2022 International,* (Reference Number: 389).

Gupta, N. & Chavan, S. R. (2021, December). Frequency analysis incorporating a decision support system over Mahanadi catchment in India, India, International Conference on Hydraulics, Water Resources and Coastal Engineering, SVNIT Surat, India. December 23-25, 2021, *ISH - HYDRO 2021 International,* (Reference Number: 357).

Gupta, N. & Chavan, S. (2021, April). Spatio-temporal characterization of rainfall using an innovative trend and discrete wavelet transformation approaches in Bhakra catchment, India. In *EGU General Assembly Conference Abstracts* (pp. EGU21-15857)**.**

Gupta, N. & Chavan, S. (2020, May). Characterization of Extreme precipitation over India. In *EGU General Assembly Conference Abstracts* (p. 12883).

Gupta, N. & Chavan, S. R. (2018, December). Assessment of Tail Behavior of Probability Distributions of Daily Precipitation Data Over India, International Conference on Hydraulics, Water Resources and Coastal Engineering, NIT Patna, India. December 19-21, 2018, *ISH - HYDRO 2018 International,* (Reference Number: 101).

Book Chapters

Gupta, N. & Chavan, S. R. (2023). Frequency Analysis Incorporating a Decision Support System Over Mahanadi Catchment in India. *In Flood Forecasting and Hydraulic Structures*. Springer Nature Singapore. http://dx.doi.org/10.1007/978-981-99-1890-4_22.

Gupta, N. & Chavan, S. R. (2021). Assessment of Tail Behavior of Probability Distributions of Daily Precipitation Data over India. *In Climate Change Impacts on Water Resources*. Springer, Cham. 10.1007/978-3-030-64202-0_10.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND NOTATIONS

AO	Arctic Oscillation
AL	Aleutian low
AES	Annual Exceedance series
BM	Block maxima
BMCMC	Bayesian Markov Chain Monte Carlo
CAES	Concentration adjusted expected shortfall
CCCR	Centre for Climate Change Research
CDF	Cumulative distribution function
CORDEX-SA	Coordinated Regional Climate Downscaling Experiment-South Asia
CI	confidence interval
CP	Concentration Profile
CM	Concentration Map
CRU	Climatic Research Unit
CMIP3	Coupled Model Intercomparison Project Phase 3
CMIP6	Coupled Model Intercomparison Project Phase 6
DSS	Decision Support System
DMR	Discriminant moment ratio plot
EPF	Exceedance Probability Function
ENSO	El Niño–Southern Oscillation
EP	Elemental Percentile
ES	Expected Shortfall
EASM	East Asian Summer Monsoon
EAWM	East Asian Winter Monsoon

ETCCDI	Expert Team on Climate Change Detection and Indices
G	Gamma distribution
GEV	Generalised Extreme Value
GP\GPD	Generalized Pareto Distribution
GPCC	Global Precipitation Climatology Centre
GCM	General Circulation Model
HIB	Halphen Type IB
IMD	Indian Meteorological Department
ISMR	Indian summer monsoon rainfall
IPCC	Intergovernmental Panel on Climate Change
IITM	Indian Institute of Tropical Meteorology
JB	Jarque-Bera
KS	Kolmogorov–Smirnov
KG	Köppen-Geiger Climate Classification
LN	lognormal distribution
LS	Least Square
LP3	Log Pearson type III distribution
LM	L-Moments
MAMP	mean annual maximum precipitation
MEF	Mean Excess Function
MS	Maximum-to-sum
MPS	Maximum Product of Spacings
MDPDE	Minimum Density Power Divergence Estimator
ML	Maximum Likelihood
OB	Obesity index

MATLAB	Matrix Laboratory
MOM	Method of Moments
MEF	Mean excess function
NAO	North Atlantic Oscillation
NFIP	National Flood Insurance Program
NMST	New model selection test
OB	Obesity Index
PDF	Probability Density Function
PDO	Pacific decadal oscillation
POT	Peak-over-threshold
PRMSE	Probability Ratio Mean Square Error
PMP	Probable Maximum Precipitation
P3	Pearson type III distribution
PII	Pareto-type II
RCP	Representative Concentration Pathways
RHBN	Reference Hydrometric Basin Network
RCM	Regional climate model
SSP	Shared Socioeconomic Pathway
SREX	Special Report on Extremes
UTR	Upper Tail Ratio
VaR	Value-at-Risk
W	Weibull
WPSH	Western Pacific subtropical high
α	Shape Parameter
β	Scale Parameter

σ	Standard deviation of discharge
KS	Kolmogorov-Smirnov
$RX1day$	Maximum 1-day precipitation
$R95p$	Very wet day
$R99p$	Extremely wet day
$L_\lambda(x)$	Lorenz Curve
λ	Truncation level
n	Sample length/ size
k	Sufficient number of observations left in the right tail
$\{G(\lambda)\}_{\lambda\in[0,1]}$	Concentration Profile
$G(0)$	Starting point of Concentration Profile
$CAES_\lambda$	Concentration Adjusted Expected Shortfall at trucncation level λ
r_1	First risk driver
r_2	Second risk driver
$R(r_1, r_2)$	Risk Function
a	Parameter of Cobb-Douglas Risk Function
b	Parameter of Cobb-Douglas Risk Function
$\bar{F}(x)$	Theoretical exceedance probability
$\bar{F}_N(x_i)$	Empirical probability of exceedance
N	Number of years of record
OB	Obesity Index
Max_count	Maximum number of times for which random sampling can be performed
X	Random variable
p_0	Probability of dry day
m	Meter
$r(x_i)$	Rank of the precipitation equal to x_i

n_d	Average number of days in a year
x_L	Threshold extracted from the non-zero daily precipitation data
x_p	Lower threshold value corresponding to probability p
$e(x_p)$	Mean Excess Function
$F_X^{-1}(p)$	quantile function
F_X	probability distribution function
Count	Number of optimum sampling
z_stat	Z statistics
R	Set of real number
C_v	Coefficient of Variation
C_s	Coefficient of Skewness
$C_v, C_s / (\hat{\gamma}_2, \hat{\gamma}_3)$	Couple point
$Z(u)$	Zenga Curve
$R_n(p)$	Maximum-to-sum ratio
M_n^p	Partial maximum
S_n^p	Partial sum
H	Elevation

1.1   Overview

Precipitation is a climatic phenomenon with high spatial and temporal variability. It is treated as a continuous random variable bounded at zero and has a Probability Density Function (PDF) comprising a decreasing monotonic tail. Precipitation extremes generally lie in the upper tail part of the frequency distribution. Often, extreme precipitation events tend to occur rarely, making it difficult to predict their magnitudes for higher return periods (i.e., for smaller exceedance probabilities). Several studies have acknowledged the shortcomings of exponentially tailed PDFs for representing the daily extreme precipitation (e.g., Koutsoyiannis, 2004a, 2004b; Panorska et al., 2007; Papalexiou and Koutsoyiannis, 2013; Papalexiou et al., 2013; Serinaldi and Kilsby, 2014; Cavanaugh et al., 2015; Beskow et al., 2015; Papalexiou and Koutsoyiannis, 2016; Papalexiou et al., 2018). The conventional distribution fitting methods cannot adequately fit the tail of daily precipitation data, which results in the exemption of extreme precipitation events. Identifying the best-suited probability distribution is a matter of eminent significance, as the wrong selection may either overestimate or underestimate the design precipitation for hydraulic infrastructure, causing detrimental consequences like floods (e.g., Koutsoyiannis, 2004a, 2004b; Panorska et al., 2007; Li et al., 2012; Papalexiou and Koutsoyiannis, 2013; Chen and Brissette, 2014; Serinaldi and Kilsby, 2014; Zhanling et al., 2015; Papalexiou and Koutsoyiannis, 2016; Papalexiou et al., 2018). There is a growing accord that precipitation extremes can be suitably represented by heavy-tailed distributions, which are an alternative to the exponentially tailed PDFs (Wilson and Toumi, 2005; Strupczewski et al., 2011; Papalexiou et al., 2013; Cavanaugh et al., 2015; Beck et al., 2017; Wietzke et al., 2020; Moccia et al., 2021; Nerantzaki and Papalexiou, 2022). A heavy-tailed distribution is one where the upper tails decay as a power law (i.e., tails tend to approach zero less rapidly than an exponential tail). These distributions are also referred as "fat-tailed", "thick-tailed", or "long-tailed" according to various literature (El Adlouni et al., 2008; Foss et al., 2013; Papalexiou et al., 2013; Nerantzaki and Papalexiou, 2019). A heavy-tail distribution implies that the extremes are more likely than those predicted by distributions like Exponential, Gamma, and Gumbel, having exponential asymptotic behavior (El Adlouni et al., 2008). Commonly considered heavy-tailed distributions are the Pareto Type II distribution, Kappa distribution, Generalized Logistic distribution, Cauchy distribution, Fréchet distribution, Weibull distribution with shape parameter $\alpha < 1$, Lognormal distribution, and so on (Foss et al., 2013; Pinheiro and Ferrari, 2015; Panahi, 2016). Papalexiou et al. (2013, 2018) intuitively defined two broad classes of distributions based on the asymptotic property of tails, viz., (a) sub-exponential class (heavy-tailed) with tails decreasing more slowly than any exponential tail and (b)

super-exponential class or hyper-exponential class (Vela and Rodríguez, 2014) having tails approaching zero more rapidly than the exponential tail.

Several other studies are related to the classification or grouping of probability distribution tails according to their general properties and limiting behavior (Ouarda et al., 1994; Werner and Upper, 2004; El Adlouni et al., 2008). These have reported the existence of five nested classes of distributions ordered from heavy to light as $A \subset B \subset C \subset D \subset E$ (see Figure 1.1), here stable distributions lie in class A, Pareto-type tail distributions in class B, Regularly varying distributions in class C, Subexponential distributions in class D, and Exponential distributions in class E. Further, these studies have highlighted the importance of selecting a class of distributions that provide the best fit to a dataset, especially for the right tail, before selecting a particular model (El Adlouni et al., 2008; Wietzke et al., 2020).
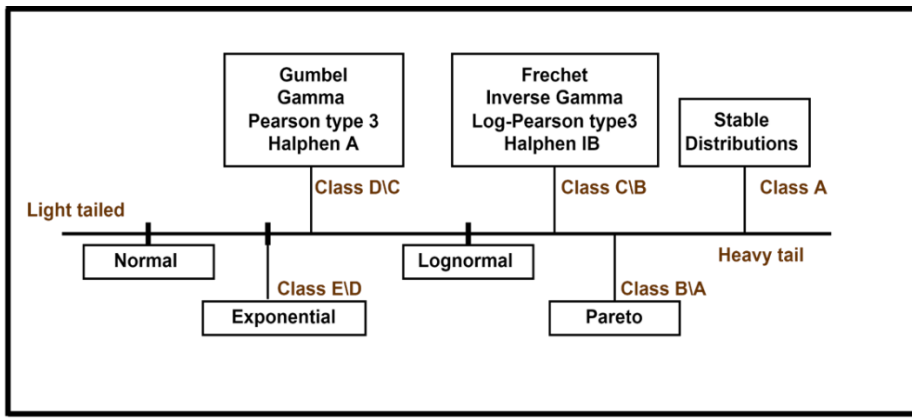
Figure 1.1 Distributions ordered with respect to their upper tails. Distributions are ordered from light-tailed (from the left) to heavy-tailed (to the right (see El Adlouni et al., 2008). Given that all these classes are nested, the class of distributions that belong to class C2 and not to class C1, such that, will be noted as C2\C1. For example, the class of sub-exponential distributions that are not regularly varying is noted as D\C.

Characterization of tail behavior helps in understanding the likelihood of the occurrence of extremes. In the wake of the inability of simple parametric models to fit the whole range of data equally well, hydrologists, climatologists, and statisticians have developed new and advanced approaches for distribution fitting required for quantile estimation in such a way that they attach greater weight to the tails of distributions than to the whole range of data to help with better characterization of the tails of the probability distributions of the hydroclimatological datasets (Coles et al., 2003; Cirillo, 2013; Fontanari et al., 2018a; Nerantzaki and Paplexiou, 2019; Nerantzaki and Papalexiou, 2021). However, the quantification and assessment of upper tail behavior (i.e., tail heaviness) are not straightforward (Cooke et al., 2014). Three main ways that are generally employed to do so include (1) Threshold-based approaches (like PMSRE, Bayesian) (Papalexiou et al., 2013; Moccia et al., 2021), (2) Quantitative or Scalar diagnostic tools (shape parameter of typical distributions (e.g., GEV, GP), obesity index (OB), Gini index, Upper tail ratio (UTR), and surprise factor) (Nieboer, 2011; Cooke et al., 2014; Sartori and

Schiavo, 2015, Smith et al., 2018; Wietzke et al., 2020), (3) Graphical interpretation (Mean excess plots, log-log plots, Hill ratio plot) (Hill, 1975; El Adlouni et al., 2008; Ghosh and Resnick, 2010; Das and Ghosh, 2016; Langousis et al., 2016; Konapala et al., 2017; Fontanari et al., 2018a, b, Nerantzaki and Papalexiou, 2019).

1.2   Motivation

1.2.1   Need to characterize the daily precipitation datasets over India using threshold-based approaches (Papalexiou et al., 2013, Nerantzaki and Papalexiou, 2019)

The threshold-based approach provides inferences associated with the frequency and magnitude of extreme events by segregating the extreme events from the precipitation data and fitting a suitable probability distribution function to them. As per the classical extreme value theory, the block maxima (BM) extracted from a time series resemble one of the three limiting distributions, namely, (i) Gumbel distribution (i.e., Extreme Value Type I distribution); (ii) Fréchet distribution (i.e., Extreme ValueType II distribution); and (iii) reversed Weibull (i.e., ExtremeValue Type III distribution) (Fisher and Tippett, 1928; Gnedenko, 1943, Jenkinson, 1955; Coles et al., 2001; Langousis et al., 2016). In the case of annual maximum daily precipitation (i.e., BM per year), Gumbel and Fréchet distributions were found to be appropriate to model the behavior of extremes as both possess unbounded upper tail behavior (Koutsoyiannis, 2004; Papalexiou and Koutsoyiannis, 2013; De Michele, 2019). The inferences from the BM approach are observed to be dependent on the selection of block size (i.e., either annual or seasonal maxima, etc.). The selection of the annual maxima (AM) from daily precipitation records at a location may distort the tail behavior of their probability distribution as it might miss a few of the largest daily precipitation events from a particular year. As the BM approach discards a large portion of information from available data, the estimated distribution parameter exhibits significant variability and becomes sensitive to outliers (Coles et al., 2003; Langousis et al., 2016). Another way of modeling extreme precipitation is based on the peak-over-threshold (POT) approach. In the POT approach, a sample is extracted from a daily precipitation series by selecting all observations above an arbitrary threshold u (Chow, 1964). As the threshold increases, such samples tend to follow generalized Pareto distribution (GPD) (e.g., Balkema and de Haan, 1974; Pickands, 1975). Many studies from the past revealed that the findings based on the POT approach are generally more efficient than the BM approach (Cunnane, 1973; Caires, 2009; Villarini et al., 2011; Moccia et al., 2019). Despite its advantages, the use of POT is less prevalent than BM due to (i) the presence of serial dependence in identified peaks and (ii) ambiguity in the selection of an optimum threshold for the identification of peaks (e.g., Beguería, 2005; Mailhot et al., 2013; Serinaldi and Kilsby, 2014; Kiran and Srinivas, 2021).

To avoid the loss of information about the extreme precipitation, as in the case of the BM approach and the selection of arbitrary threshold in the case of the POT approach, an annual exceedance

series (AES) can be used for demarcating the tail of daily precipitation data. An AES approach has the advantage of better representing the exact tail of the parent distribution. After finalizing the tail of daily precipitation data, the distribution fitting can be accomplished using the probability ratio mean square error (PRMSE) norm proposed by Papalexiou et al. (2013). PRMSE norm-based tail-fitting of the probability distributions yields unbiased estimates of the parameters of distributions and also compares the fits of various probability distributions (e.g., Papalexiou et al., 2013, 2018; Moccia et al., 2019). The discussion above signifies the need to apply the AES-based approach to find the best-suited distribution based on the PRMSE norm that can describe the extreme daily precipitation in India.

1.2.2 Need for the development of an empirical index that can provide a quantitative comparison between two distributions and alleviate the shortcomings associated with the threshold-based approaches

Diagnostics that are based on the graphical interpretation of distributions utilize, for instance, mean excess plots, log-log plots, or the generalized Hill ratio plot (Hill, 1975; Ghosh and Resnick, 2010; Nieboer, 2011; Cooke et al., 2014; Das and Ghosh, 2016; Roth et al., 2016) for assessing the tail behavior of distributions. These methods are sensitive to the selection of a threshold for segregating the tail part of the distribution and need visual interpretation for assessing tail behavior, making the graphical methods more time-consuming. Also, these graphical methods are usually restricted to certain classes of distributions; hence, a quantitative comparison of tail heaviness between two or more distributions belonging to different classes is difficult. Although there are recent attempts to make graphical tools like the mean excess function more objective (Nerantzaki and Papalexiou, 2019), yet graphical methods are hardly feasible when a high number of samples are to be compared. Hence, there is a need for quantitative or scalar, objective, and easily applicable indicators. Cooke and Nieboer (2011) proposed the concept of the obesity index (OB) to diagnose the behavior of the distribution tail. The OB is a quantitative, scalar, and easy-to-use indicator that measures tail heaviness (Wietzke et al., 2020). The concept of the OB was proposed based on the heuristic that a heavy-tailed sample usually has a few large values compared to the other values (Cooke and Nieboer, 2011; Cooke et al., 2014). The concept overcomes the limitations associated with graphical or threshold-based approaches (Wietzke et al., 2020). The major advantage associated with the OB is that it does not require assuming any threshold for segregating the tail part of a distribution (i.e., it can measure the heaviness of the distribution tail without referring to the limiting behavior of a distribution) (Nieboer, 2011; Cookeet al., 2014). This approach checks how far the largest sample values occur from the main body of the probability distribution, which is also known as the principle of "a single big jump". This principle can form a basis for determining the presence of sub-exponentiality in the data. Furthermore, the index is an empirical measure and can be computed directly from the data without assuming any form of distribution function. The above discussion signifies the need to explore and develop the OB index further and to assess its

utility for performing a diagnosis and characterizing the heaviness of tails of daily gridded precipitation data over India.

1.2.3 *Need to develop a Comprehensive Decision Support System (DSS) to characterize the tails of the dataset by alleviating the limitation of Conventional DSSs*

The scalar indices are definitely easy to use, but their application is restricted only to identifying the behavior of the tail as light or heavy. For example, an OB > 0.75 for a data set indicates the presence of a heavy tail (i.e., heavier than exponential tail) but does not identify the suitable distribution class (i.e., the family of distributions) or model (i.e., distribution). The importance of identifying the class of distributions that provides the best fit to the upper tail of a dataset before selecting a particular model has been highlighted by many researchers like El Adlouni et al. (2008), Ehsanzadeh et al. (2010) and Martel et al. (2013). A practical approach to select the appropriate class of distribution for a dataset considering its right tail (i.e., extreme) is termed as Decision support system (DSS). The conventional DSS proposed by El Adlouni et al. (2008) and Ehsanzadeh et al. (2010) utilizes various graphical approaches such as MEF plot, log-log plot, and the generalized Hill ratio plot to characterize the tails of distributions (Hill, 1975; Beirlant et al., 2004; Ghosh and Resnick, 2010; Nieboer, 2011; Cooke et al., 2014; Roth et al., 2016; Das and Ghosh, 2016; Nerantzaki and Papalexiou, 2019). However, the conventional DSSs have some shortcomings like (i) consideration of LN as a part of class D due to more conservative results (i.e., overestimation) in the quantile estimation study and (ii) unavailability of suitable criteria to discriminate amongst the classes C, D, and LN (Ehsanzadeh et al., 2010; Martel et al., 2013). This discussion suggests the need for the development of a comprehensive DSS that can alleviate the shortcomings of the conventional DSS. Overall, there is a dearth of attempts to incorporate some advanced graphical methods into the conventional DSS to improve the characterization of tails of datasets.

1.2.4 *Need for a framework to assess the relative risk associated with the tails of the daily precipitation*

Extreme climatic events, such as extreme rainfall and temperature, profoundly impact human lives and society. The pattern of extreme precipitation has undergone substantial changes worldwide caused due to global warming, and these modifications are evident in terms of changes in both the frequency and magnitude of precipitation (Westra et al., 2014). Understanding of the altered frequencies and magnitude can be examined based on the assessment of the tails of the probability distributions. Identification of the upper tail behavior, as well as quantification of the associated risk, becomes important for risk mitigation. While numerous ways are present to characterize the tail of distribution for daily precipitation, there is a notable lack of dialogue regarding the associated risks linked with these tails (i.e., extreme events). Value-at-Risk (VaR) and Expected Shortfall (ES) are the common risk measures; however, they are not preferred when the tails of the datasets tend to follow skewed heavy-

tailed distribution due to their inability to measure the actual risk dispersed in the tail. Several concentration measures, such as the Lorenz Curve and Gini index, can efficiently represent the risk for the skewed distributions through the construction of new risk indicators. These measures can effectively investigate and identify the relevant facts as well as statistical regularities of the datasets based on some analytical and geometrical properties (related to the Lorenz curve). The efficacy of the new risk indicators to quantify the tail risk of skewed distributions has not been tested yet for characterizing the hydroclimatic datasets. Hence, this discussion highlights the need for the development of a framework to assess the relative risk associated with the tails of distributions of daily precipitation over India.

Overall, recognizing and appropriately accounting for heavy tails in precipitation datasets is crucial for several reasons. Firstly, it can help in the assessment of the likelihood of rare and severe precipitation events to equip us against the potential risks linked to floods and other hydrological hazards. Secondly, changes in the tail behavior of precipitation distributions can provide insight into the shifts in extreme event frequency and intensity under changing climate. This understanding is vital for assessing climate change impacts and devising effective adaptation strategies.

1.3   Objective of the study

The following objectives have been set for this thesis based on the motivations listed in the foregoing section.

1. To characterize the daily precipitation dataset over India using a threshold-based approach and assess the temporal change in the tail behavior due to climatic shift in the 1970s.
2. To characterize the tail behavior of daily precipitation probability distributions over India using the Empirical Indices.
3. To develop a comprehensive Decision Support System for the characterization of probability distribution tails for daily precipitation.
4. To develop a framework for assessing the embedded risk associated with the tails of the daily precipitation over India.

1.4   Thesis Outline

A chapter-wise breakup and the summary of the proposed thesis work as follows:

Chapter 1: Introduction

An introductory background about the precipitation, extreme events, heavy and light tail distributions, implications of underestimating tails, Classes of distributions, and different ways to assess the tail heaviness are discussed to provide a broader outlook on the research area. The introduction is

followed by stating the motivation behind carrying out the present work along with the objectives of this thesis.

Chapter 2: Literature Review

A comprehensive literature review on the fundamentals of tails, techniques or approaches used for the characterization of tails, and assessment of embedded tail risk in the context of this thesis, along with the details of the literature related to extremes under changing climate over India, has been discussed. Various important studies done by researchers over the past few decades reporting the presence of heavy tails in the precipitation datasets are put forward. Several innovative techniques used worldwide for assessing tail heaviness are reviewed and analyzed for the scope of further development of the proposed research area.

Chapter 3: Assessment of temporal change in the tails of the probability distribution of daily precipitation over India due to climatic shift in the 1970s

This chapter presents a conventional threshold-based approach given by Papalexiou et al. (2013) that makes use of the annual exceedance series and is used to determine the best-suited distribution based on the Probability Ratio Mean Square Error (PRMSE) norm. The approach is applied to the Indian daily gridded precipitation dataset, and an attempt is made to assess the tail behavior of the daily precipitation dataset and the temporal change in the tail behavior over India from pre-to post-1970 time periods as per the global climatic shift at the grid level and regional level. The assessment of temporal changes in magnitude and frequency of extreme precipitation due to climatic shifts is also taken to produce the severity maps over India.

Chapter 4: Characterizing the tail behavior of daily precipitation probability distributions over India using the Obesity Index

This chapter proposes an algorithmic approach based on a novel scalar upper tail indicator known as the 'Obesity Index', that can quantitatively diagnose the heaviness of distribution tails without assuming any threshold for segregating the tails, thereby alleviating the limitation of previous threshold-based approaches. The step-by-step algorithm for determining OB using the bootstrapping technique helpful for reliable discrimination of various distribution tails is provided. A simulation study recognizing the behavior of the OB for the variation in the shape and scale parameter of various distributions is presented by generating samples from four commonly used probability distribution functions, namely, Pareto, Weibull, Lognormal, and Gamma. Investigation related to adequate sample length and the optimum number of random samples required for the application of the algorithm is also presented. Investigation into the sensitivity of precipitation quantiles to the change in OB corresponding to various non-exceedance probabilities is also undertaken. In this study, the OB approach was used to discern the probability distribution of the daily gridded precipitation data for historical and future

periods over India into light and heavy tails. Also, a comparative analysis between the OB-based approach and threshold-based approach by Papalexiou et al. (2013) and Nerantzaki and Papalexiou (2019) is provided to examine their ability to characterize the probability distributions of the daily gridded precipitation over India into two broad classes of distribution that are, subexponential and exponential/hyper-exponential. The OB-based approach is extended to examine the change in tail behavior over time by considering various climate change scenarios from the Coordinated Regional Climate Downscaling Experiment-South Asia (CORDEX-SA). Exploration of the dependence of the obesity index on the amount of the annual maximum daily precipitation and the elevation of grids across India is also shown in the chapter. The presents the spatial variation of OB with respect to the elevation across India and its relation with the climatology of the country. Different results grouped for each of the subdivisions of India are also presented.

Chapter 5: *A comprehensive Decision Support System (DSS) for the characterization of probability distribution tails for daily precipitation*

In this chapter, a novel algorithmic framework of a comprehensive DSS useful for selecting the appropriate class of distribution, specifically focusing on the right tail of a dataset, is presented. The proposed DSS incorporates advanced graphical methods like concentration profile (CP) plot, Concentration adjusted expected shortfall (CAES) plot, Zenga plot, maximum-to-sum plot, and Discriminant Moment ratio plot (DMR) plot, in a precise manner, either as an identification test or a confirmatory test to categorizes different classes of distributions like classes B\A (Pareto type), C\B (regularly varying), D\C (subexponential), E (Exponential type), hyper exponential class (outside class E) and LN (Lognormal) distribution (the limiting case between class C and D), thereby alleviates the shortcomings associated with the conventional DSS. The robustness of the proposed DSS over the conventional approaches is established through a simulation study where sample data of different lengths from representative probability distributions belonging to various classes (e.g., D, C\B, B\A, E, etc.) and limiting case LN are generated and reclassified successfully into their respective classes. Furthermore, the chapter presents the evaluation of the influence of sample length on the effective implementation of the DSS. Finally, the utility of the proposed DSS is demonstrated through its application to daily gridded precipitation data over India.

Chapter 6: *Assessment of embedded risk in precipitation tails over India through Concentration Profiles: A multi-model assessment from CMIP6 experiments*

In this Chapter, inferences from tools like '*Concentration Profile*', which are used to estimate the tail variability of any distribution, are incorporated into a risk function to construct a new risk measure, which can directly determine the risk associated with the tails of different probability distributions. A 2-dimensional map, known as a '*Concentration Map*' (CM), is utilized to assess the riskiness of different datasets. A simulation experiment is presented, which helps in understanding the nature and variation

of risk based on the risk function parameter values. Consequently, the utility of the risk function is demonstrated through its application to estimate the riskiness of the daily gridded precipitation datasets over India for historical and future time periods. The approach is used for the spatiotemporal variation analysis by examining the change in risk patterns over India for historical as well as future time periods based on the simulations from various climate change scenarios from Coupled Model Intercomparison Project Phase 6 (CMIP6). Furthermore, a regional scale investigation is also presented that helps understand the variation of tail risk within 34 meteorological subdivisions and climate zones based on climate classification schemes like Köppen-Geiger (KG) over India during different periods.

Chapter 7: Summary and Conclusions

This Chapter will provide significant conclusions from this thesis work and some inputs on the future scope of work possible in this area of research.

2.1   Overview

Precipitation extremes can have significant and widespread adverse impacts on both the environment and human society. These impacts include flooding, erosion, landslides, water quality issues, agricultural impacts, inefficient water resource management for major infrastructure, infrastructure damage, economic costs, and ecological consequences. Extreme precipitation events are usually found in the upper part (often referred to as the 'tail') of the probability distribution function (PDF). They occur infrequently, which presents challenges in predicting their magnitudes for higher return periods. Numerous studies have recognized the limitations of exponentially-tailed probability density functions (PDFs) for modeling daily extreme precipitation. As an alternative, many researchers suggest the use of heavy-tailed distributions to represent daily precipitation extremes. Heavy-tailed phenomena are commonly observed in the field of hydrology and are important to be detected and assessed, as most risk reduction measures are based on the probability of extreme events. Numerous advances have been made in ways of characterizing the tails of probability distributions and examining the risk associated with these tails using various approaches.

The subsequent sections in this chapter provide the necessary definitions and formulations needed for understanding the concept of heavy and light-tailed distributions, along with the details related to the concept of classes of distributions. The ultimate goal of past research has been to introduce various methods for accurately diagnosing the heavy-tailed behavior of hydroclimatic datasets like precipitation. This chapter reviews some early work related to the characterization of tail behavior using different approaches around the world. Additionally, it presents literature works that provide intrinsic details related to the different approaches like threshold-based approaches, empirical or quantitative indexes, and graphical approaches, which are essential for providing context and motivation as mentioned in Chapter 1. The chapter also includes an overview of the early work and commonly accepted definition of the Decision support system and Risk assessment tools. Furthermore, this Chapter discusses details related to climate extremes in India, potential causative factors, and the effect of global climatic shifts on extreme precipitation. It also delves into literature focusing on the impact of climate change on future extremes in India. The chapter provides different conclusions with an outlook that explains the outputs of the literature review in the context of the thesis objectives.

2.2   Fundamentals on tail characterization of various classes of distributions

Various probability distributions can be classified based on their tail behavior into two categories: heavy-tailed distributions (e.g., Pareto, Lognormal, Weibull, Lévy, etc.) and light-tailed distributions (e.g., Gaussian, Exponential, etc.). Defining what constitutes a heavy or light tail is a topic of common discussion, and many studies have concluded that there is no universally accepted definition. Ambiguity persists as heavy-tailed distributions are also referred to as "fat-tailed", "thick-tailed," or "long-tailed" in various literature (El Adlouni et al., 2008; Foss et al., 2013; Papalexiou et al., 2013; Weitzke et al., 2020). Most often the heavy tails are defined as the ones that have upper tails decaying as a power-law (i.e., tails tend to approach zero more gently than an exponential tail). A random variable X is said to have a heavy tail when its moment-generating function becomes infinite on R (the set of real numbers) given in Equation 2.1 (e.g., Bryson, 1974; Mikosch, 1999; Panorska et al., 2007; Foss et al., 2013; Panahi 2016; Wang et al., 2018).

$$\int_R e^{-\lambda x} F(x) dx = \infty \quad \text{for all} \quad \lambda > 0 \tag{2.1}$$

Some studies define heavy tail behavior as the power-law behavior of the upper tail. The exponential distribution on the other hand is given by the PDF

$$p(x) = \lambda e^{-\lambda x}, \qquad x \geq 0, \quad \lambda > 0 \tag{2.2}$$

which is often considered as the boundary between classes of heavy-tailed and light-tailed distributions. For light-tailed distributions, all moments exist and are finite.

Typically, the heavy-tailed distributions comprise the class of sub-exponential distributions. The class of sub-exponential distribution was initially introduced by Chistyakov (1964), and for any distribution function F to be subexponential, one of the following conditions must hold.

(a) $\lim_{x \to \infty} \dfrac{\bar{F}^{n*}(x)}{\bar{F}(x)} = n$ \qquad for some (all) n $\geq$ 2, \hfill (2.3)

(b) $\lim_{x \to \infty} \dfrac{P(X_1 + ... + X_n > x)}{P(\max(X_1, ..., X_n) > x)} = 1$ \quad for some (all) n $\geq$ 2, \hfill (2.4)

where $\bar{F}(x)$ denotes the exceedance probability; $\bar{F}^{n*} = 1 - F^{n*}(x) = P(X_1 + X_2 + ... + X_n > x)$ denotes the tail of n-fold convolution of F (Embrechts et al., 1997; Goldie and klüppelberg, 1998; Embrechts and Goldie, 1982). Definition (a) shows the absence of any exponential moments, while condition (b) suggested by Teguel (1975) indicates that the sum of n independent and identically distributed (iid) subexponential distributions are likely to exceed x if and only if its maximum value is larger x (in simple word it shows the presence of enormous values or rare events in the sample). Condition (b) is also known as the principle of "a single big jump" (Foss et al., 2013; Hill, 2019). Embrechts and Goldie (1980)

showed the equivalence of conditions (a) and (b). Power-law distributions like Lognormal distribution and regularly varying distributions such as Pareto type II distribution are sub-sets of the sub-exponential distributions (Feller, 1971; Bingham et al., 1987; El Adouni et al., 2008; Foss et al., 2013; Voitalov et al., 2018).

Several attempts have been made to group tails of distributions according to their limiting/asymptotic behavior (e.g., Goldie and Klüppelberg, 1998; Ouarda et al., 1994, etc.). Werner and Upper (2004) classified the distributions in five nested classes from A to E such that $A \subset B \subset C \subset D \subset E$. Here, Class A includes stable distributions (distributions with Pareto tails having $\alpha < 1$), Class B includes Pareto-type tails, Class C includes regularly varying distributions, Class D includes sub-exponential distributions, and Class E includes exponential distributions. Class E represents the broadest class whose upper tails decrease exponentially (or more slowly) and, thus, more slowly than the tails of normal distributions. They are characterized by the tail behavior presented in Equation 2.2. Class D distributions are the ones given by the conditions in Equations 2.3 and 2.4. They have tails that decrease slower than the exponential tail. Going one step further, Class C is applied as a limiting class. The main characteristic of distributions in this class is that, far out in the upper tail, the tails decrease similar to that of the Pareto distribution as shown in Equation 2.5 given below.

$$\lim_{t \to \infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha} \tag{2.5}$$

The exponent or parameter α is called as the ''tail index'' and is used as a measure of the tail heaviness. Distributions in class B have exact Pareto tails. The cumulative distribution function of the Pareto distribution is

$$F(x) = 1 - u^\alpha x^{-\alpha} = 1 - \left(\frac{u}{x}\right)^\alpha, \quad x \geq u \text{ and } u > 0 \tag{2.6}$$

The tail index α can be related to the moments of a distribution with Pareto tails. Indeed, the probability density function of a random variable following the Pareto distribution is $f_{Pareto}(x) = \alpha u^\alpha x^{-\alpha-1}$ and the k-moment is given by

$$E[X^k] = \alpha u^\alpha \int_u^\infty x^{k-\alpha-1} dx \tag{2.7}$$

Thus, only k-moments such that $k < \alpha$, are finite. This property is important to define the last class (class A). Note that small values of α imply heavier tails. Classes C and B are very important considering their connection to classical extreme value theory (Koutsoyiannis, 2004; El Adouni et al., 2008; Ehsanzadeh et al., 2010). Furthermore, Class A contains stable (or α-stable) distributions. Stable

distributions have a Pareto tail with $\alpha < 2$, which implies infinite variance and, as a consequence, very fat tails.

El Adouni et al. (2008) combined the classifications mentioned above with 5 graphical criteria for tail discrimination and arranged them from light to heavy-tailed where E being the lightest and A being the heaviest. Papalexiou et al. (2013, 2018) intuitively defined two broad classes of distributions based on the asymptotic tail behavior as (a) the sub-exponential class (heavy-tailed class), and (b) the superexponential class (Nagaev and Tsitsiashvili, 2006) or hyper-exponential class (Vela and Rodríguez, 2014) (light-tailed class). The practical implication of a heavy-tailed distribution such as a Pareto or a lognormal distribution is that the large values representing rare events are much more likely to occur than that of a light-tailed distribution like Gaussian or Exponential distribution. Hence, it becomes important for us to understand these tails in detail and depth.

2.3   Characterization of tail behavior of probability distributions

This section of the chapter discussed qualitative as well as quantitative techniques used to characterize the extremes or tails (in general). There are many methods or approaches that allow us to estimate the characteristics of extreme events.

2.3.1   Conventional or Threshold based approaches

Traditionally, data-driven methods for selecting appropriate probability density functions (PDFs) for daily rainfall have largely involved fitting specific PDFs to data at limited locations, as demonstrated by Katz et al. (2002). Many researchers have proposed the characterization or classification problems of heavy *vs.* exponential tail problems in the context of threshold-based approaches like Block Maximas (or Annual Maxima) (BM) (or AM) and Peaks over threshold (POT) methodology. The BM (or AM) method consists of selecting maximum rainfall values, one for each year of observation. The POT sample is achieved by selecting all extreme values that exceed an arbitrarily fixed threshold (Chow, 1964). Several authors provide a comparative survey on the adaptation of different theoretical probability distributions to hydrological empirical samples selected by these two approaches. Madsen et al. (1997a, 1997b) propose a regional estimation scheme for extreme events modeling by selecting AM and POT samples for discharge data recorded in New Zealand. Their findings reveal that the POT-Generalized Pareto Distribution (GPD) model is generally more efficient than the AM-Generalized Extreme Value (GEV) model. Panorska et al. (2007) were among the pioneers in investigating high-frequency precipitation extremes within the context of climate variability and change. They used the results derived from probability theory to develop an efficient automated scheme to distinguish between heavy and exponential precipitation probability density function (PDF) tails in hundreds of daily station records spanning five decades over the North American continent. They approached the heavy *vs.* exponential tail classification problem in the context of peaks over threshold (POT) methodology and derived a new statistical test based on the theory of maximum likelihood ratios. They verified that the

daily extreme precipitation data at most stations in North America comes from a Pareto distribution or resembles power law rather than an exponential distribution. They concluded that traditional statistical distributions (e.g., exponential, Weibull, Gamma, lognormal) used for modeling daily rainfall generally underestimate extreme probabilities, with the extent of distortion, or volatility, dependent on regional and seasonal climatic variations. The Pareto/exponential likelihood ratio for POT emerged as a valuable tool for diagnosing tail behavior and distinguishing between exponential and heavy-tailed PDF families, although threshold selection remains largely intuitive.

Similarly, Cavanaugh et al. (2015) used the methodology of Panorska et al. (2007) along with the statistical test (likelihood ratio test of exponentiality *vs* Pareto) of Kozubowski et al. (2009) to characterize the distributions of daily precipitation exceedances over suitably high thresholds for gauge-based station records on six continents over the globe. The probability tail structure of over 22,000 weather stations globally is examined in order to identify the physically and mathematically consistent distribution type for modeling the probability of intense daily precipitation and extremes. Results showed the dominance of the Pareto-type tail over 65% of stations as compared to the exponential tail-type distributions. This implies that statistical distributions most often used to model daily rainfall (e.g., exponential, Weibull, Gamma, and lognormal) generally underestimate the probabilities of extremes. They also found that the magnitudes of these discrepancies, i.e., volatility, depend on seasonal and regional climate characteristics over the globe. Heavy tails are most prominent in regions that experience high-valued precipitation from many different types of weather events that produce wildly different precipitation rates similar to Panorska et al. (2007).

The application of the POT approach necessitates determining the optimal threshold, a challenge acknowledged by many researchers (Koutsoyiannis, 2004a; Begueria, 2005; Mailhot et al., 2013). Apart from threshold selection methods, another difficulty of the POT approach is the assumption of data series independence (Bezak et al., 2014). To address these issues, Papalexiou et al. (2013) introduced the annual exceedance series (AES) approach, a threshold-based method that tackles the challenges of conventional POT by eliminating the arbitrary selection of the threshold. Instead, it offers a way to indirectly determine the threshold based on the empirical distribution, ensuring that the number of values above the threshold matches the number of years (N) in the record (refer to, for example, Cunnane, 1973). They then applied this approach in conjunction with a modified mean square error (MSE) norm known as the Probability Ratio Mean Square Error (PRMSE) norm. The approach facilitates the direct fitting and comparison of different theoretical distribution tails to the empirical tails estimated from nearly 15,137 daily rainfall records worldwide. Theoretical tails, such as those of the Pareto, Lognormal, Weibull, and Gamma distributions, are fitted to the empirical ones by minimizing the numerically PRMSE norm. The distributions were ranked from best to worst in describing the tails of the daily precipitation based on the PRMSE norm. The overall findings of the study suggested that heavy-tailed distributions provide a better description of daily rainfall extremes at most of the stations. Serinaldi and Kilsby (2014) used two worldwide data sets and performed a detailed investigation on how the threshold

selection and record length affect the right tail behavior of POT observations. They focused on the shape parameter of the Generalized Pareto distribution and identified two main effects: (1) as the threshold decreases the Generalized Pareto shape parameter variance reduces converging to positive values, yet the asymptotic Generalized Pareto hypothesis becomes less realistic tending to be replaced by Weibull stretched exponential tails, and (2) given a fixed high threshold and increasing record length the variance of the Generalized Pareto shape parameter decreases with its mean value converging to positive values.

Further, Papalexiou et al. (2018) presented the climatology of the tail and quantified its heaviness in over 4,000 hourly precipitation records across the United States using a novel Bayesian adjustment approach. They compared two major types of tails based on their vast popularity across many scientific fields, that is, power-type, or else Pareto tails and Weibull tails which include stretched-exponential, exponential, and hyper-exponential tails. They defined the empirical tails using POT definitions using a fixed number of peaks, for example, *m* largest values in an *m*-year sample (similar to Papalexiou et al., 2013), and investigated the impact that empirical tail have on tail fitting using PRMSE norm. They advocate the use of the PRMSE norm for comparing the theoretical and empirical tails as it uses relative errors between theoretical and empirical values, and thus, each point contributing to the sum is equally weighted. PRMSE can be considered a straightforward method to determine the best-fit performance between CDFs since the distribution that has the lowest RMSE value is the one that guarantees the best adaptation to the empirical sample. They stressed that identifying correctly the type of tail is not trivial and assessed the precision of the tail-fitting method using Monte Carlo (MC) simulations. They explored which distribution tail between power-type and Weibull tails better describes hourly precipitation extremes based on the tail indexes. Both the PII and W tail-index values showed that hourly precipitation has a heavy (subexponential) tail, much heavier than exponential or Gamma tails that would significantly underestimate precipitation over large return periods. The spatial variation of both tail indices shows a coherent pattern over the United States, with mountainous areas exhibiting heavier tails. They revealed a nonlinear increase in the tail's heaviness with elevation and provided corresponding parametric functions to describe this law. Precipitation in mountainous areas was fitted especially by the heavier tails.

Moccia et al. (2021) also investigated the tail behavior of the daily precipitation records for two Italian regions, namely Lazio and Sicily, located in central and south Italy, respectively. They employed two coupled methods: (i) PRMSE (Papalexiou et al., 2013) and (ii) the Kolmogorov-Smirnov test (KS; Keutelian, 1991) to assess the best-fitting probability distribution for samples selected with the AM and AES, respectively. Specifically, they tested six theoretical probability distributions (i.e. Gamma (G), Lognormal (LN), Weibull (W), Gumbel (Gu), Frechet (F), and Pareto type II (P)) to fit the empirical daily rainfall data recorded in two Italian regions. They found that for the two regional datasets (i.e. Sicily and Lazio), the Frechet distribution provides the best fit for the AM samples, while AES samples are optimally fitted by the Pareto type II. The PRMSE and the KS test were shown to be equivalent in terms of results. However, the PRMSE offers two main advantages as it allows the comparison and the

definition of the best-fitting distribution and its parameters. The difference in terms of return period and the related precipitation values is notable, for each of the six candidate distributions, and it increases dramatically for high return periods, especially for the AM samples. They showed that the sample selection method of extreme rainfall values leads to significantly different results and more research is needed to further investigate this aspect as in the engineering practice, the AM is the most widely adopted method. Finally, Rajulapati et al. (2021) assessed the tail heaviness using five global gridded precipitation products. They fitted the Pareto and Weibull tails and found discrepancies among the different products, strong spatial variability in tail heaviness, and significant differences in return levels for large return periods between the two tails. In the context of assessing the tail behavior of precipitation datasets, explorations of the tail behavior of daily precipitation datasets belonging to Indian landmass datasets are definitely needed using threshold-based approaches.

2.3.2 *Empirical (quantitative) Approaches*

The choice between heavy-tailed and exponentially-tailed models is qualitative in nature, and the approaches mentioned in section 2.3.1 are among the few popularly used methods to address the problem of characterizing or classifying different tail types, such as heavy or exponential tails. In addition to these approaches, there are a few other diagnostics based on the graphical interpretation of distributions that utilize tools such as MEF plot, log-log (or Zipf) plot, or the generalized Hill ratio plot for assessing the tail behavior of distributions (Zipf, 1949; Hill, 1975; Stanley et al., 1995; Kratz and Resnick 1996; De Sousa and Michailidis, 2004; Ghosh and Resnick, 2010; Nieboer, 2011; Cooke et al., 2014; Das and Ghosh, 2016; Roth et al., 2016). The MEF gives the expected excess of a random variable over a certain threshold given that this random variable is larger than the threshold. The log-log function is essentially a plot of the empirical survivor function on logarithmic axes, involving the fitting of a straight line above a specified threshold. The slope of this line is then used to estimate the tail index. Another method, the Hill plot, is based on the idea that if a random variable has a Pareto distribution, then the log of this random variable has an exponential distribution with a parameter equal to the tail index. The Hill estimator is an estimator of the parameter of this exponential distribution. The use of these methods in different fields like hydrology, actuarial science, survival analysis, environmental science, and economics has been acknowledged by many researchers (Resnik, 2007; Das and Resnick, 2008; Ghosh and Resnick, 2010; El Adlouni et al., 2008; Panahi, 2016). However, there are certain drawbacks associated with these methods which are identified by several researchers (Cirillo, 2013; Xie, 2017). Methods like MEF and log-log are sensitive to the selection of a threshold for segregating the tail part of the distribution and need visual interpretation for assessing tail behavior, making them time-consuming. In the case of the MEF, it is also seen that if we consider a regularly varying distribution function with a tail index ($\alpha < 1$), then the excess function of this distribution function does not exist. However, when plotting the empirical mean excess function, the slope of this plot remains finite. The Hill plot estimator faces challenges related to large variance and bias, depending on the number of

largest observations taken into account to estimate the tail index (Caeiro et al., 2005; Danielsson et al., 2019; Németh and Zempléni, 2020). Also, Another limitation associated with the Hill ratio plot is that it works very well for Pareto-distributed data, but for other regularly varying distribution functions, it becomes less effective (Nieboer, 2011). Considering the limitations of individual graphical methods, it is evident that these techniques are typically constrained to specific classes of distributions, making it challenging to quantitatively compare the tail heaviness of distributions from different classes. The lack of objectivity of individual graphical methods necessitated the need for quantitative or scalar, objective, and easily applicable indicators (Nieboer, 2011; Cooke et al., 2014)

Earlier research utilized various scalar upper tail indicators to investigate the upper tail behavior of hydroclimatic or economical variables. Upon reviewing the literature, four pertinent scalar upper tail indicators were identified: shape parameter (or tail index), Upper Tail Ratio (UTR), Gini Index, and Obesity Index (OB). Shape parameters or tail index of different distributions have been used by several researchers to quantify the upper tail behavior of flood and heavy precipitation distributions (Bernardara et al., 2008; Smith, 2010; Papalexiou et al., 2013; Gu et al., 2017; Zhou et al., 2017; Papalexiou et al., 2018). Hobbi (2021), explores the changes in the frequency and intensity of extreme precipitation for different climate types (Köppen-Geiger's climate classification) using over 8582 daily station rainfall records from the Global Historical Climatology Network-Daily (GHCN-D) database. The author estimated the magnitude and significance of trends of the annual maximum precipitation time series by applying non-parametric tests of Mann-Kendall and Sen's slope estimator and measured the heaviness of the tail based on the shape parameter of the fitted Generalized extreme value (GEV) distribution estimate using the L-moments (Hosking, 1990). The findings of the study reveal a significant increasing trend in 9.7% of stations in the eastern USA, Asia, and northern Europe. However, only 2% of stations in eastern Australia and central USA had a significant decreasing trend. Largest to smallest heavy-tailed extremes in major climate types E (polar), A (tropical), B (dry), D (snow), and C (temperate). For climate subtypes, large heavy-tailed extremes were observed in Dfd, ET, and Am, while only light-tailed extremes were observed in Cfc. It has been seen that using shape parameter or tail index as a tool for characterization of the heavy-tailedness of a distribution is quite common but can have some drawbacks like, sometimes estimation of the tail index is difficult from a dataset as it can only be observed at infinity and not directly from a dataset (Ghosh, 2016).

Further, Villarini and Smith (2010) and Villarini et al. (2011b) introduced the UTR, calculated as the highest value in the sample normalized by the 10-year return level. It is a nonparametric approach to analyze the upper tail of distributions. Smith et al. (2018) utilized UTR to examine the annual peak observations from more than 8,000 U.S. Geological Survey (USGS) stream-gaging stations and found that the nature of flood peak distributions in the conterminous United States is unbounded and thick-tailed. In the intervening time, the Gini index, a traditional inequality measure from economics, has also been recently proposed as another upper-tail indicator (Eliazar and Sokolov, 2010; Fontanari et al. 2018a, 2018b). Gini is intuitively derived from the Lorenz curve by most researchers like Kondor et al.

(2014), and Konapala et al. (2017), however, the literature also mentions some other ways to determine it (Eliazar and Sokolov, (2010)). Originally rooted in economic contexts, the Gini index has made its way into hydro-meteorological sciences, proving useful for capturing inequality and temporal changes in the distributions of daily precipitation (Rajah et al., 2014; Lai et al., 2018), streamflow (Masaki et al., 2014; Zhang et al., 2015), and river solute loads (Jawitz and Mitchell, 2011). Rajah et al. (2014) assess changes in the temporal uniformity (or lack thereof) of wet-day precipitation amounts using the Gini index for a data set of 12,513 land-based stations from the Global Historical Climatology Network. In the context of rainfall, the Gini index emerges as a crucial indicator of the uneven distribution of precipitation throughout a year, displaying values on a scale from 0 (indicating a uniform distribution across all days) to 1 (representing concentrated precipitation on a single day). Distinguishing itself from alternative measures of variability, such as standard deviation, the Gini index proves to be robust and dimensionless, facilitating clear interpretation across various geographical settings. Notably, the Gini index does not directly correlate with tail heaviness. Instead, it provides a quantification of the impact of both small and large values within a distribution (Wietzke et al., 2020).

Another scalar indicator, known as the Obesity Index was introduced by Nieboer (2011) which can be used to diagnose the behavior of the distribution tail. Nieboer (2011) found OB to be an effective quantitative, scalar, and easy-to-use indicator that measures tail heaviness. They characterized the data on flood insurance claims (National Flood Insurance Program (NFIP), crop loss claims (National crop losses of the U.S.), hospital discharge bills (U.S. Hospital Discharge Bills), precipitation (based on Geographically based economic data, i.e., G-Econ database (Nordhaus et al., 2006)) and damage and fatalities from natural catastrophes (SHELDUS database) by using the concept of OB. They also pointed to the need for future research to investigate under which conditions the Obesity index increases as the tail index of a regularly varying distribution decreases. Sartori and Schiavo (2015) also applied the concept of OB for the investigation of the upper tail behavior of negative shocks in global agricultural production. Hill (2019) presented the Cooke-Nieboer index based on the concept of OB and used it to distinguish the networks in real life drawn from the ICON database (Clauset et al., 2016), having power-law, exponential, and symmetric degree distributions.

Although several upper-tail indicators coexisted, yet a comparative analysis was not present till recent times. This gap was filled by Wietzke et al. (2020) who conducted a comparative study in which they evaluated the performance of the shape parameter (GEV), obesity index, Gini index, and upper tail ratio against a novel benchmark of tail heaviness – the surprise factor. They found that UTR replicates the surprise factor best but is most uncertain and only comparable between records of similar length. For samples with symmetric Lorenz curves, shape parameters, obesity, and Gini indices provide consistent indications. They suggest the use of a combination of shape parameters, obesity, and the Gini index to characterize tail heaviness. From the literature review, various scalar indicators and their properties and applications in the past have been identified. However, in the pursuit of further exploration and development of scalar indicators tailored for hydrological variables, it is noteworthy

that the scalar indicator, the Obesity Index (OB), could be further enhanced and employed to assess the heaviness of tails in precipitation datasets. Remarkably, the OB index has seen limited application in the realm of hydrological extremes and presents an opportunity for further development and utilization in this context.

2.3.3   *Graphical Approaches*

In the literature review, various methods for tail analysis are identified, including the threshold-based approach, graphical approach, and empirical (or scalar indicator) approaches. Among graphical and scalar indicators, the latter are easy to use; however, their application is restricted to identifying the behavior of the tail as light or heavy. For example, an Obesity Index (OB) > 0.75 for a dataset indicates the presence of a heavy tail (i.e., heavier than an exponential tail) but does not identify the suitable distribution class (i.e., the family of distributions) or model (i.e., distribution). Conventional graphical methods also lack the ability to diagnose multiple types of tails as mentioned in section 2.3.2; however, certain advancements have introduced better graphical tools that can characterize tails belonging to different classes simultaneously. Cirillo (2013) introduced a novel graphical plot known as the Discriminant moment ratio plot (DMR), which is nothing but the simpler and extended version of the CV-Skewness diagram (Vargo et al., 2010) can be efficient enough to scrutinize the inferences revealed from the log–log, and MEF plots. Cirillo (2013) also proposed the Zenga plot, which helps to discriminate between the lognormal and the Pareto distributions, two classes that are difficult to identify as both tails have similar representations on graphical plots. Zenga curve can be expressed analytically for any distribution via the corresponding Lorenz curve ($L(u)$) (Lorenz, 1905). The Zenga plot shows a Zenga curve ($Z(u)$) plotted against u (threshold varying between 0 and 1) and assumes different shapes for different distributions. They found that conventional graphical tools like MEF or log-log plots are not always reliable and only a combination of these tools can give some degree of confidence about the real presence of different tail types. The additional tools proposed by them showed better efficacy in refining the tail assessment or characterization analysis.

Later, Cirillo and Taleb (2016) used a combination of different graphical tools to examine statistical pictures of violent conflicts over the last 2000 years. They make use of a novel approach to deal with fat-tailed random variables with a remote but nonetheless finite upper bound, by defining a corresponding unbounded dual distribution (given that potential war casualties are bounded by the world population). They analyzed the presence of a long right tail (or Paretian tail) in the distribution of victims using several graphical plots like Q-Q plot, MEF plot, and another interesting graphical tool which is the maximum-to-sum plot (MS plot or R_n^p). MS ratio is nothing but a ratio of partial maximum (M_n^p) and partial sum (S_n^p) for p statistical moments (Herein $p = 1$: mean; $p = 2$: variance; $p = 3$: skewness; and $p = 4$: kurtosis). The power-law classes (Class B and Class C) are characterized by the non-existence of higher-order moments, while distributions belonging to LN and class D\C have the existence of all

their moments. The MS Plot relies on a simple consequence of the law of large numbers (Embrechts, 2003; Cirillo and Taleb, 2020; Manz and Mansmann, 2020) and sees the non-convergence or convergence of R_n^p to zero for inferences about power-tail or non-power-tail type distributions, respectively. Their data analysis suggests a heavy right tail for the distribution of war casualties, both for raw and rescaled data. They recommended using such approaches in other fields of science where power laws play a role in modeling, like geology, hydrology, statistical physics, and finance. Greselin et al. (2017) also applied Zenga's inequality curve to assess the progression, redistributive effects, and re-ranking effects of a personal income tax system. The Zenga curve, comparing economic conditions across population percentiles, offers unique insights distinct from traditional Lorenz curves. Cirillo and Taleb (2020) also analyzed data for pandemic outbreaks spanning over the past 2500 years using different graphical tools in combination and found that the related distribution of fatalities is strongly fat-tailed, suggesting a tail risk that is unfortunately largely ignored in common epidemiological models. Graphical tools enabled them to conclude that the distribution of the victims of pandemic diseases might be a distribution with no finite moments. Another advanced graphical approach was introduced by Fontanari et al. (2018a) known as concentration profile (CP) plots and Concentration adjusted expected shortfall (CAES) plots, which can identify the different parametric families of the loss distribution, especially Lognormal, Pareto, Exponential, and Weibull ($\alpha > 1$ and $\alpha < 1$), simultaneously. These plots were based on concentration (or inequality) measures (e.g., the Gini index derived from the Lorenz curve and common risk management measures like the Expected Shortfall (ES) (Acerbi and Tasche, 2002; McNeil et al., 2015; Cirillo and Taleb, 2016). Fontanari et al. (2018a) preferred the sequence of truncated Gini indices (indicated by $G(\lambda)$, where λ denotes truncation level) over the conventional Gini index, as it measures the dispersion above the Value at Risk (VaR) (Jorion, 2001) to have a reliable measure of tail risk and precision of the ES. They showed that the financial data deviate from Gaussianity and often exhibit heavy-tailed behavior meaning that rare and disruptive events have a non-negligible chance of happening.

MEF, a conventional graphical tool, was enhanced by Das and Ghosh (2016), who constructed confidence intervals for MEF plots to ascertain the domain of attraction (Fréchet, Gumbel, or reversed Weibull) of the data. They applied their methodology to simulated and observed datasets, including regional ozone concentrations and flow rates. The literature also mentions recent efforts made by Nerantzaki and Papalexiou (2019) to develop a faster algorithmic procedure for MEF to discriminate between exponential and sub-exponential tails. They were the first to apply the algorithmic approach to nearly 21,348 daily precipitation records worldwide to assess whether the observed data can be described by exponential or heavier tails. They observed that nearly 75.8% of records showed the dominance of heavy-tail distributions across the globe, except for eastern South America. Lighter tails are evident in regions such as eastern South America (eastern Brazil, Uruguay, Paraguay), specific areas in central Africa, and central China. In contrast, heavier tails are observed in the eastern U.S., central

Canada, parts of India, and South Africa. The heaviest tails are found in Australia and Eurasia, with a notable "hot spot" located in central Russia and Kazakhstan. Despite the effectiveness of the approach, it did not resolve the issue of identifying and differentiating the tails from LN, class C/D, and class D/E. Also, this approach has limitations associated with confidence interval and sample size availability.

Researchers such as El Adlouni et al. (2008), Ehsanzadeh (2011), Cirillo (2013), and Martel et al. (2013) have emphasized the preference for using a combination of graphical tools rather than a single one, as it enhances the identification of the appropriate distribution class. They have underscored the importance of determining the class of distributions that best fits the upper tail of a dataset before selecting a specific model. A practical approach for selecting the appropriate distribution class for a dataset, particularly considering its right tail (i.e., extreme), is referred to as a Decision Support System (DSS). A detailed discussion of various aspects of DSS is provided in the subsequent subsection. Regarding graphical tools, the literature suggests their robustness in assessing the tail behavior of distributions. However, there is a lack of attempts to integrate these advanced graphical methods together to enhance the characterization of dataset tails. Additionally, the application of these graphical methods for hydrological variables could open new avenues for research.
.

2.4   *Decision support system (DSS)*

A DSS utilizes various graphical methods to characterize the tails of distributions as per their limiting behavior and groups them into the appropriate class of distributions. Popular classes of the probability distributions are given by Ouarda et al. (1994) and Werner and Upper (2004) in a nested form as $A \subset B \subset C \subset D \subset E$, starting from light-tailed distributions to heavy-tailed distributions) (see Fig. 1.1 in Chapter 1). Five nested classes of distributions include stable distributions in class A, Pareto-type tail distributions in class B, Regularly varying distributions in class C, Subexponential distributions in class D, and Exponential distributions in class E. Details for which are already mentioned in section 2.2. Literature reveals different versions of the DSS, the one of the conventional DSS was initially proposed by El Adlouni et al. (2008). El Adlouni et al. (2008) proposed to make use of a set of graphical criteria that are developed in the extreme value theory framework (Embrechts et al. 2003) to select the class of distributions that seems to adequately represent the sample extremes. They showed the application of the DSS to assess the tail of the observed annual peak flow time series of the Potomac River for the time period 1895-1986 and 1895-2000, respectively. Later, Ehsanzadeh et al. (2010) assessed the usefulness of the proposed DSS to characterize the tails behavior of sample data belonging to three large hydroclimatic databases, Hydrometric Basin Network (RHBN), precipitation, and UNESCO. They confirm that discriminating between classes of competing models prior to model selection is critical when the sample data comes from extreme events. They recommended the use of a normality test to detect the log-transformed data as LN3 is a transformation of the normal distribution. Their study highlights that choosing an inappropriate class of distributions for model selection often results in

underestimating the quantity of the variable under study. This underestimation poses a higher socioeconomic risk compared to the risk associated with overestimating a specific quantile. Further, Martel et al. (2013) advanced the conventional DSS by incorporating some steps prior to the conventional DSS. They developed a LN3 goodness-of-fit procedure, based on the coefficient of variation, the coefficient of skewness, and the Jarque-Bera normality test. Their study actually presented an effective way of discriminating between the LN3 distribution and that of Class C of regularly varying distributions (heavier tail) and D of subexponential distributions (lighter tail). In conclusion, while Decision Support Systems (DSS) appear to be effective tools for tail categorizations, existing literature lacks more refined DSS structures and definitions utilizing advanced tools. Furthermore, there is a noticeable absence of efforts to integrate advanced graphical methods into DSS, which could significantly enhance the characterization of tails of the hydrological datasets.

2.5   *Understanding the Tail Risk*

Tail risk (low-probability extreme events) technically refers to the risk associated with both the left and right tails. People working in the field of finance and economy are mostly concerned with losses (the left tail), while hydrologists or climatologists are generally concerned with extremes belonging to the right tail. The heavy-tailed distributions have much larger high percentiles relative to the rest of the data values than the exponentially-tailed ones. That implies that in places where heavy-tailed models are appropriate, and especially if observed and modeled trends continue, future large events may be much larger than those observed to date. Identification of the upper tail behavior as well as quantification of the associated risk becomes important as the majority of risk reduction measures are based on the probability of extreme events. In the past several studies have shown the use of different tools like the Lorenz Curve (Lorenz, 1905), Gini indices (Gini, 1912), and different indices associated with them to efficiently represent a risk, or to construct new risk measures, and to investigate relevant facts and statistical regularities of the datasets based on some analytical and geometrical properties of the tools (Jones and Zitikis, 2003; Methni et al., 2014; Brazauskas and Kleefeld, 2015; Furman et al., 2017; Fontanari, 2019; Loffredo et al., 2021; Chen and Cheng, 2022). The Lorenz curve is a transformation of a positive valued random variable that maps its quantile function into an increasing convex function space and the study of the geometry of such transformation can measure or capture different aspects of the variability of a random variable. The Gini index properly summarizes the information on the variability embedded in the Lorenz curve yet a single Gini value is not sufficient to describe the entire behavior of the Lorenz curve (Arnold, 2012; Eliazar, 2018).

Value-at-Risk (VaR) and Expected Shortfall (ES), the common risk measure, are not preferred by many researchers in the case when the tails of the datasets tend to follow skewed heavy-tailed distribution as they fail to measure actual risk dispersed in the tail for such cases (Linsmeier and Pearson, 2000; Jorion, 2001; Hull, 2012; Embrechts et al., 2013; McNeil et al., 2015; Cirillo and Taleb, 2016).

The Gini index, which is quite rooted in economics has found its way to hydro-meteorological sciences and was applied to capture inequality and temporal changes of distributions of daily precipitation, streamflow, and river solute loads as mentioned earlier. Fontanari et al. (2018a) define the truncated version of the Gini index, as a function of the VaR and denoted it as a Concentration Profile (CP). They combined CPs and standard results from utility theory to develop a tool known as a Concentration Map (CM) that assesses the riskiness of datasets taking into account the variability of the larger and most relevant events. Overall, they presented an important way of dealing with the issue of tail variability measurements for portfolio loss distributions using the descriptive power of the Lorenz curve. In the context of studies assessing the embedded tail risk of hydro-meteorological variables, such as daily precipitation datasets, there is a scarcity of research. Conducting these studies is crucial in the current scenario where changes are inevitable, and appropriate methods need to be applied.

2.6   *Climate Extremes over India (Past and Future)*

India recognized as a hotspot for global warming due to its diverse and intricate geography (Krishnan et al., 2020; Huang et al., 2022), experiences a highly varied climate influenced by numerous factors (Rajeevan et al., 2012; Niranjan Kumar et al., 2013). The southwest monsoon, contributing approximately three-fourths of the annual rainfall, plays a crucial role in sustaining the Indian subcontinent, supporting millions of people, and forming a cornerstone of the economy (Kumar et al., 2005; Kumar et al., 2010; Yadav, 2013). Analysis of historical data reveals a relatively stable pattern of rainfall over the past century. However, the escalating global average temperature has led to an increase in extreme events (Goswami et al., 2006). While inter-annual variations in monsoon rainfall constitute only about 10% of the long-term average, the occurrence of extreme rainfall events significantly contributes to the occurrence of floods and droughts, resulting in substantial economic and human losses. It is well documented that variability of seasonal rainfall over India is associated with sea surface anomalies in the tropical Pacific Ocean, such as those related to El Niño and Southern Oscillation (ENSO), in the Indian Ocean, such as the Indian Ocean Dipole (IOD) and the Arabian sea (Rajeevan et al., 2012; Azad and Rajeevan, 2016). The changes in extreme precipitation can also be attributed to the abrupt global change of the climatic system caused by a regime shift in the 1970s in various climatic factors like the Arctic Oscillation (AO), East Asian summer monsoon (EASM), East Asian winter monsoon (EAWM), North Atlantic Oscillation (NAO), Aleutian low (AL), Western Pacific subtropical high (WPSH) and Indian summer monsoon rainfall (ISMR) (Biondi et al., 2001; Chowdary et al., 2006; Chowdary et al., 2006; Zhou et al., 2009; O'Kane et al., 2014; Chen et al., 2015a; Sahana et al., 2015; Hsu et al., 2016; Zuo et al., 2016; Weisheimer et al., 2017; Dai et al., 2018) or some local changes such as urbanization. These agents, in combination or independently affect the moisture transport dynamics, which in turn induces variability in precipitation patterns. In the past, the climate regime shift has adversely impacted the atmosphere, ecosystems, biological and many hydro-climatic variables, such as temperature, air pressure, wind field, and rainfall resulting in the frequent occurrence of extremes like

heat, drought, heavy rainfall, and flood disasters (Graham, 1994; Zhang et al., 1998, Wang, 2001; Meehl et al., 2008; Jacques-coper and Garreaud, 2015; Huang et al., 2017a, b).

There have been several other studies that analyzed the changes in extreme rainfall over India at the national and regional scale and have drawn different conclusions (Goswami et al., 2006; Rajeevan et al., 2008; Dash et al., 2009; Kulkarni et al., 2012; Ghosh et al., 2012; Shastri et al., 2015; Mondal and Majumdar, 2015; Singh et al., 2016; Ghosh et al., 2016; Roxy et al., 2017; Bisht et al., 2018a, 2018b). Many researchers have demonstrated that the rainfall characteristics in the past over India, within and beyond the monsoon period exhibit spatio-temporal changes due to climate regime shifts globally and in India (Sabeerali et al., 2012; Sahana et al., 2015). Over central India, Goswami et al. (2006) showed an increasing trend in both the magnitude and frequency of extreme precipitation, likely to result in severe property damage in the future. Ajayamohan and Rao (2008) have also shown an increased extreme rainfall event over central India after the 1976/1977 climate shift. Rajeevan et al. (2008) used 104 years (1901–2004) of high-resolution daily gridded rainfall data and reported significant variabilities in inter-annual and inter-decadal changes in addition to a statistically significant long-term trend in the frequency of extreme rainfall events. Vittal et al. (2013) showed that rainfall extremes changed in India after 1975 and established that urbanization, in terms of change in population density, is a possible cause of change. They used a comprehensive POT approach with 95 and 99th percentile thresholds, including multiple extreme events in a year. Roxy et al. (2017) found that there has been a threefold increase in extreme rain events over India during 1950-2015. This indicates that extreme rainfall events have increased over time, and there is an urgent need to investigate the dynamics of their occurrence and their role in augmenting risks. Dash and Maity (2019) found that the precipitation-based climate change indices exhibit increasing trends over India with more spatial extent post-1975. Recently, Sarkar and Maity (2020) observed an increment of 35% in Probable maximum precipitation over India in the post-1970 (1971–2010) period as compared to the pre-1970 (1901–1970) period due to climate shift. Heavy rainfall extremes could be a result of external forcings which include both human-induced and natural activities or/and internal forcings such as internal mechanisms within the climate system (IPCC, 2014). However, there exists substantial debate in the recent literature about the spatio-temporal distribution of extreme rainfall events over India and their relationship with various aspects of global climate change. For instance, Kulkarni (2012) suggested that there is a rising trend in rainfall extremes co-occurring along with decreasing moderate rainfall causing an insignificant overall trend. However, percentile-based frequency and intensity analysis of extreme rainfall events showed no visible spatially uniform trends over India (Ghosh et al., 2012). These conflicting conclusions about extreme rainfall events in current research point out the necessity of a comprehensive evaluation of rainfall extremes.

The country has also witnessed noticeable changes in climatic conditions marked by increased floods and droughts during the past few decades especially post-1970s (post-industrial era), which are likely to persist in the future. Interseasonal, Interannual, and Interdecadal climatic variations in India have a significant relationship with various external forcings. Simulating the historical and current

climatic conditions of India is, therefore, a difficult undertaking. It is also essential to forecast future changes in precipitation extremes so that local governments can implement adaptation and mitigation strategies before time. Future climate change relies heavily on reliable and accurate projections. CMIP6 uses the new Shared Socio-economic Pathways (SSP) scenarios, enabling us to observe the probable changes in precipitation extremes worldwide and over different regions like India (Eyring et al., 2016; O'Neill et al., 2016). The latest generation of climate models has shown significant advancements to their previous version (CMIP3 or CMIP5) in reproducing large-scale spatiotemporal patterns of climatic variables. These improvements are in terms of finer horizontal resolution, enhanced cloud microphysics parameterizations, better representation of synoptic processes, improved agreement with the global energy balance, and increased participation of modeling groups (O'Neil et al., 2016; Eyring et al., 2019). These enhancements have been acknowledged by many researchers in the arena of hydrology and climate change (Gusain, 2020; Li et al., 2021; Kamruzzaman et al., 2021; Choudhary et al., 2021; Supharatid et al., 2021; Dutta and Maity, 2022). Many studies have assessed the variation of precipitation extremes in India under some scenarios of CMIP5 and CMIP6. Maity et al. (2016) predicted an increase in precipitation in the future, with a major increment in southern cities. They observed inconsistencies among different CMIP5 models. Mukherjee et al. (2017) found that the frequency of precipitation extremes is projected to rise more prominently in southern and central India in the middle and end of the 21st century under the representative concentration pathway (RCP) 8.5. Yaduvanshi et al. (2021) explore the potential changes in the Expert Team on Climate Change Detection and Indices of rainfall and temperature estimated from the coupled model inter-comparison project CMIP5 multi-model ensemble over different climatic zones of India at 1 ℃, 1.5 ℃, 2 ℃, 2.5 ℃ and 3 ℃ global temperature rise relative to pre-industrial levels under two Representative Concentration Pathways, RCP4.5 and RCP8.5. They found that the annual total precipitation and heavy rainfall-related extreme indices show statistically significant increases in the tropical, temperate, and semi-arid regions of India, moving from 1 ℃ to 3 ℃ warming level under the RCP8.5 scenario whereas there is generally no significant change in the maximum number of consecutive dry and wet days. Though several models in the recent phase (CMIP5) incorporated new components such as dynamic vegetation, indirect effects of aerosols, etc. (Taylor et al., 2012), but their coarse spatial resolution fails to capture the influence of local scale features (such as topography, land-surface feedback, land use changes, etc.) in reproducing present climatic conditions (Ghosh et al., 2016; Sharma et al., 2018; Jain et al., 2019). Several studies also pointed out the drawback of CMIP5 models to over- or underestimate the monsoon characteristics over the South Asian and Indian subcontinent inconsistently for different precipitation indices thus lowering the confidence in future projections (Saha et al., 2014; Sharmila et al., 2015). Various downscaling approaches, statistical and dynamical (Kannan and Ghosh, 2013; Salvi et al., 2013; Xue et al., 2014), were proposed in the last decade to alleviate the simulations at the local scale but they did not add any significant improvement in all cases studies and deteriorated to even worse in few (Singh et al., 2017; Sharma et al., 2018). To overcome these challenges, improved climate model simulations under the sixth phase of CMIP are released by a few modeling groups which are now used to observe

the extremes over India. Mishra et al. (2020) utilized the bias-corrected CMIP6 projections to estimate the frequency of rainfall and temperature extremes for an administrative region (state of Uttar Pradesh, India) and a river basin (Godavari, India) during near-, mid- and end-21st century and reported higher frequencies for the far period than the near-term climate. Saha and Sateesh (2022) showed that regions of Central India, North-East India, Western Ghats, and Eastern Ghats are found to be susceptible to extreme rainfall zones over the Indian landmass under three scenarios (SSP126, SSP245, SSP585) of CMIP6 models.

ASSESSMENT OF TEMPORAL CHANGE IN THE TAILS OF PROBABILITY DISTRIBUTION OF DAILY PRECIPITATION OVER INDIA DUE TO CLIMATIC SHIFT IN 1970s

3.1   Overview

The magnitude and frequency of extreme precipitation can be assessed by studying the upper tail behavior of probability distributions of daily precipitation. Depending on the tail behavior, the distributions can be classified into two categories: heavy-tailed (sub-exponential) and light-tailed (hyperexponential) distributions. Heavier tails indicate more frequent occurrences of extreme precipitation events. In this chapter, we have analyzed the temporal change in the tail behavior of daily precipitation over India from pre-to post-1970 time periods as per the global climatic shift. We intend to perform the assessment by considering 4 theoretical distributions (e.g., Pareto type II, Lognormal, Weibull, and Gamma distributions) which belong to different classes of distributions (i.e., sub-exponential or hyper-exponential classes) following Papalexiou et al. (2013). Contrary to widely used Block Maxima (BM) and Peak-over-Threshold (POT) approaches, the Annual Exceedance Series (AES) approach would be considered to demarcate the tail of daily precipitation data. A modified Probability Ratio Mean Square Error norm (PRMSE) is used to identify the best-fit distribution to the tails of daily precipitation, among four theoretical distributions as mentioned in changing climate over India. A categorical classification of grids into two broad classes of distribution, i.e., the subexponential class and the hyperexponential-exponential class is also provided with consideration of the shift in the global climatic regime in the 1970s. Further, investigation related to the spatial and temporal changes in the behavior of tails of the probability distribution of daily precipitation between the two time periods, viz. pre-1970 (1901-1970) and post-1970 (1971-2010) is studied. The assessment is performed both at grid and regional scale (i.e., Meteorological Subdivisions).

3.2   Data and Methodology

3.2.1   Description of the study area

India, the largest South Asian country with a wide variety of climatic regions extending from low rainfall arid regions to heavy rainfall receiving regions, is our study area. The climatic condition of the Indian mainland is influenced by various geographical and relief features like the Himalayas in the north, Thar Desert and Arabian Sea in west, the Bay of Bengal in the east, Western Ghats in the south-west and the Indian Ocean in the south. The study area covers a widespread range of variations in the rainfall

extremes, which motivates us to examine spatial and temporal behavior of daily extreme precipitation in terms of magnitude and frequency of occurrence between the two time periods corresponding to the shift in global climate regime in the 1970s, i.e., pre-1970 and post-1970. Further, the temporal changes are investigated at a regional scale in 34 out of 36 homogeneous Meteorological subdivisions (see Figure 3.1) in this analysis (Guhathakurta and Rajeevan, 2008).
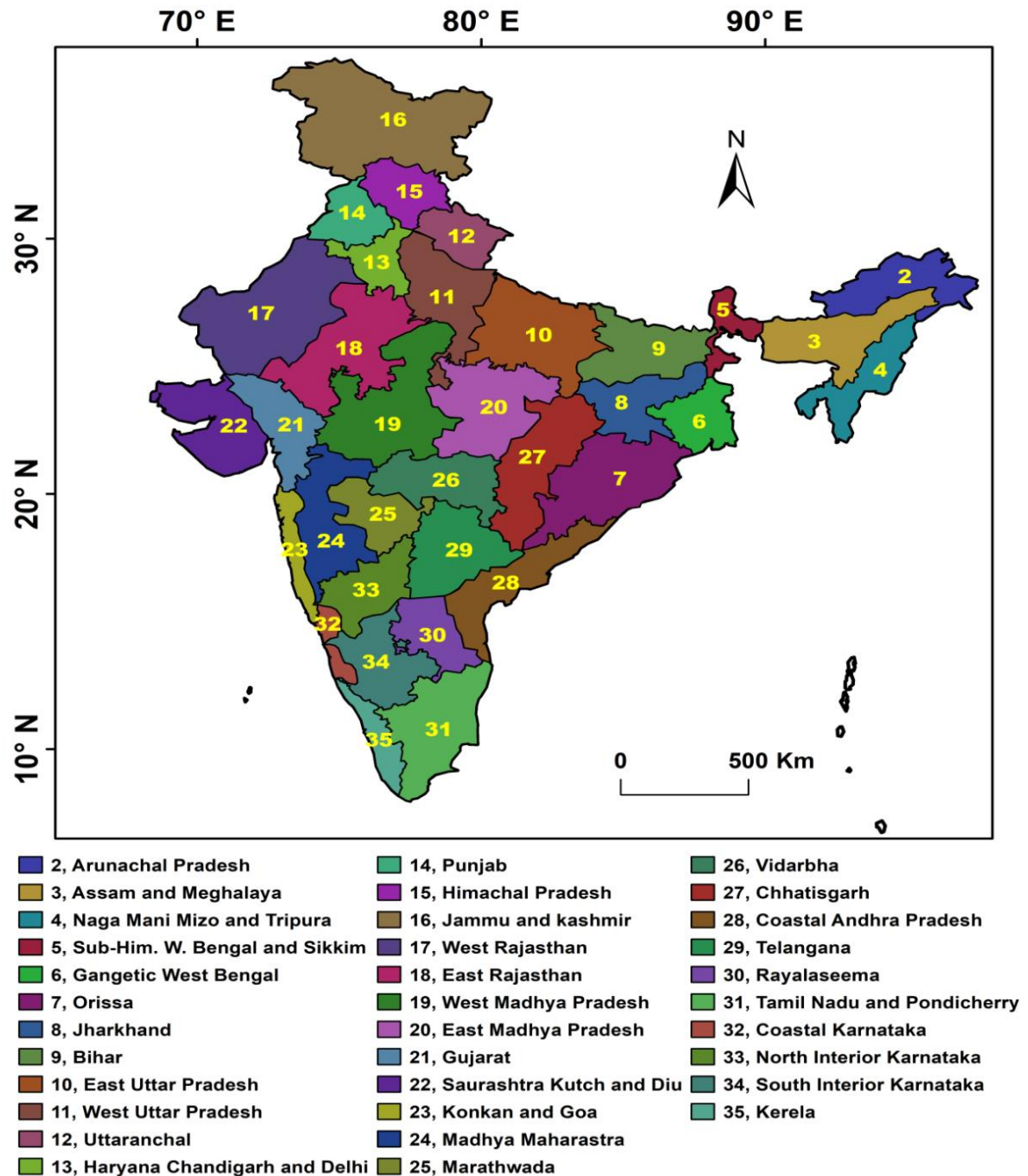
Figure 3.1 Meteorological Subdivisions of India considered for regional analysis

3.2.2   Description of the Data Used

In this study, an extensive database of daily gridded precipitation having a spatial resolution of 0.25° procured from the India Meteorological Department (IMD) is considered. The gridded rainfall data was prepared for 112 years (1901-2013) by Pai et al. (2014) using a varying network of 6955 rain gauge stations. After performing a quality check, 4789 grids each having a record length (N) of 110 years, i.e., from 1901-2010, were selected for analysis. No missing data were filled at any grids/stations. Records at each grid have been split into two parts, i.e., pre-1970 (1901-1970) and post-1970 (1971-2010), to capture the effect of the shift in the global climate regime. Despite being an unequal division of the data, the records at each grid for both pre and post-1970 time period have a sufficient number of non-zero daily precipitation values needed to estimate the tail behavior using the threshold-based approach for fitting probability distributions to daily precipitation data proposed by Papalexiou et al. (2013). The data division fulfills the condition of availability of at least a 30-year record customary in the climate community (Arguez and Vose, 2011). Individual data length for both periods is sufficient to obtain a robust representation of the spatial pattern of the tails of the probability distribution of daily precipitation data over India.

3.2.3   Threshold-Based Approach for Fitting Probability Distributions

For the analysis, we have adopted the threshold-based approach (i.e., AES) for fitting probability distributions to the tail part of the probability distribution of non-zero daily precipitation data proposed by Papalexiou et al. (2013). Since the investigation evolves around the tail behavior, it is essential first to define the part of the probability distribution known as the 'tail'. The demarcation of the tail of the empirical distribution for daily precipitation data by optimally selecting the threshold is a vital and crucial step in this approach. After demarcating the tail of the empirical distribution, fitting a theoretical probability distribution function to the daily precipitation data in the tail part can be accomplished by minimizing the difference between empirical and theoretical distributions.

3.2.3.1   Defining Tail of Empirical Distribution of Daily Precipitation

The upper or 'right' part of the empirical probability distribution function for non-zero rainfall is referred as the "tail". The choice of a threshold needed for defining a tail is recognized as a difficult and open problem of debate till date. Hence, to avoid a priori selection of threshold, we defined samples using the AES method in the present study. We choose a value x_L as a threshold such that the number of extreme precipitation events above it equals the number of years of record N (Cunnane, 1973; Ben-Zvi, 2009). N largest daily values of the record are preferred over each year's maximum value as the latter results in the distorted tail (Papalexiou and Koutsoyiannis, 2013; Papalexiou et al., 2013).

The total number of non-zero daily precipitation values at a station can be computed using $n = (1 - p_0) n_d N$ where $n_d = 365.25$ is the average number of days in a year, p_0 represents the probability of dry day. The empirical probability of exceedance $\bar{F}_N(x_i)$ is defined according to the

Weibull plotting position formula (Weibull, 1939; Makkonen, 2006) at each station having *N*-year record, and *n* number of non-zero precipitation values is defined as,

$$\overline{F}_N(x_i) = 1 - \frac{r(x_i)}{n+1}$$

(3.1)

where, $r(x_i)$ is the rank of the precipitation equal to x_i in the ordered sample as $x(1) < .... < x(n)$ of the non-zero values. Thus, the empirical tail is defined by the *N* largest non-zero precipitation values of $\overline{F}_N(x_i)$ with $n - N + 1 \leq i \leq n$. Note that the threshold value for precipitation is given as $x_L = x_{(n-N+1)}$.

3.2.3.2 *Theoretical distributions considered in this study*

Four simple, popular, and frequently used theoretical distributions such as Weibull (W), Lognormal (LN), Pareto type II (PII), and the Gamma (G) distributions are considered in this study following Papalexiou et al. (2013) and Papalexiou et al. (2018). Details on these four distributions are provided in Table 3.1. The distributions selected have two-parameter one is the scale parameter ($\beta > 0$) and the other is the shape parameter ($\alpha > 0$). The decision on the heaviness of tails of daily precipitation data is based on fitting 4 probability distributions to the precipitation data in the tail part of the empirical distribution. The distributions can be divided into sub-exponential and exponential-hyperexponential classes based on the estimates of the shape parameter, α. The former group comprises Pareto type II distribution, Lognormal distribution, and Weibull distribution with $\alpha < 1$, whereas the latter group includes Gamma distribution and Weibull distribution with $\alpha > 1$.

Table 3.1 Details about Probability density functions, Exceedance probability function, and, tail type of four commonly used distributions (Papalexiou et al., 2013)

Distribution	Probability Density Function (PDF)	Exceedance Probability Function (EPF)	Tail Type
Weibull distribution (W)	$f_W(x) = \dfrac{\alpha}{\beta}\left(\dfrac{x}{\beta}\right)^{\alpha-1}\exp\left(-\left(\dfrac{x}{\beta}\right)^{\alpha}\right)$	$\overline{F}_W(x) = \exp\left(-\left(\dfrac{x}{\beta}\right)^{\alpha}\right)$	Heavier than exponential for $\alpha < 1$ (sub-exponential) Exponential $\alpha = 1$ Lighter than the exponential for $\alpha > 1$ (hyper-exponential)
Gamma distribution (G)	$f_G(x) = \dfrac{1}{\beta\Gamma(\alpha)}\left(\dfrac{x}{\beta}\right)^{\alpha-1}\exp\left(-\dfrac{x}{\beta}\right)$	$\overline{F}_G(x) = \Gamma\left(\alpha,\dfrac{x}{\beta}\right)/\Gamma(\alpha)$	Exponential tail
Lognormal distribution (LN)	$f_{LN}(x) = \dfrac{1}{\sqrt{\pi}\alpha x}\exp\left(-\ln^2\left(\dfrac{x}{\beta}\right)^{\frac{1}{\alpha}}\right)$	$\overline{F}_{LN}(x) = \dfrac{1}{2}erfc\left(\ln\left(\dfrac{x}{\beta}\right)^{\frac{1}{\alpha}}\right)$	Heavy (approaching power type)
Pareto type II distribution (Lomax)	$f_{PII}(x) = \dfrac{1}{\beta}\left(1+\alpha\dfrac{x}{\beta}\right)^{-\frac{1}{\alpha}-1}$	$F_{pii}(x) = \left(1+\alpha\dfrac{x}{\beta}\right)^{-\frac{1}{\alpha}}$	Heavy (Power type tail)

Note: Each distribution has two parameter scale parameter ($\beta > 0$) and the shape parameter ($\alpha > 0$)

Here $\Gamma(\alpha)$ is a standard mathematical function called the gamma function and is defined by

$\Gamma(\alpha) = \int\limits_{0}^{\infty} t^{\alpha-1}e^{-t}dt$; complementary error function is denoted as, $erfc(x) = 2\pi^{-1/2}\int\limits_{0}^{\infty} e^{-t^2}dt$

3.2.3.3 Procedure to Fit Probability Distributions

The theoretical distributions are fitted to the precipitation records in the tail part of the empirical distribution by minimizing a PRMSE norm (which is an objective function) as given in Equation (3.2) (Papalexiou et al., 2013).

$$PRMSE = \frac{1}{N} \sum_{i=n-N+1}^{n} \left(\frac{\overline{F}(x_{(i)})}{\overline{F_N}(x_{(i)})} - 1 \right)^2 \tag{3.2}$$

The PRMSE norm is a function of the parameters β and α of the theoretical distributions. The norm is selected because (i) it is unbiased and suitable for sub-exponential distributions, (ii) it is easy to use and allows direct comparison of different distribution tails (iii) it gives equal weightage to each point in the tail which contributes to the sum as relative errors between theoretical and empirical values (Papalexiou et al., 2013; 2018).

In this study, the approach proposed by Papalexiou et al. (2013) is slightly modified by using a genetic algorithm (Goldberg, 1989; Michalewicz et al., 1992) for parameter estimation of the distributions. A genetic algorithm (GA) is a heuristic, stochastic, combinatorial, optimization technique based on the biological process of natural evolution (reproduction, crossover, and mutation). The heuristic is applied probabilistically to the discrete decision variables coded into binary strings. GA has been utilized effectively to minimize the PRMSE given in (Equation 3.2) in two ways: (i) by fitting theoretical distribution to entire precipitation data observed at a grid, and (ii) by fitting theoretical distribution to N largest values of precipitation at a grid. Figure 3.2 depicts the approach to fit different probability distributions, namely Lognormal, Pareto type-II, Weibull, and Gamma distributions to the precipitation events in the tail part for both pre-and post-1970 periods (i.e., 1901 to 1970 and 1971 to 2010). Grids for which the parameters change but the distribution remains unchanged are shown in Figure 3.2. It can be inferred from the figures that the first approach where distribution is fitted to entire non-zero precipitation data does not adequately describe the tail (refer to the black dashed line). On the other hand, the solid red line representing fitting of the distributions only to the events in the tail part appears to describe the tail adequately.
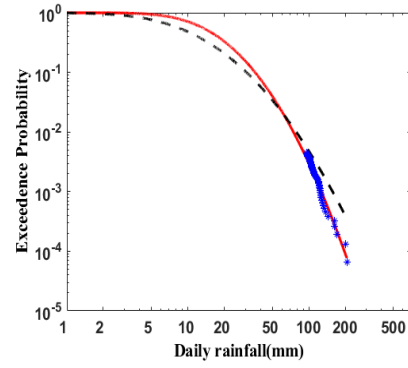
(i) Lognormal distribution (Station code: 26)
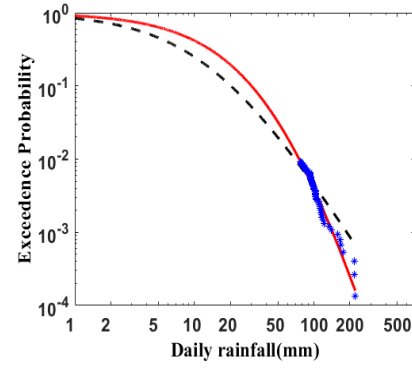
(ii) Pareto distribution (Station code: 14)
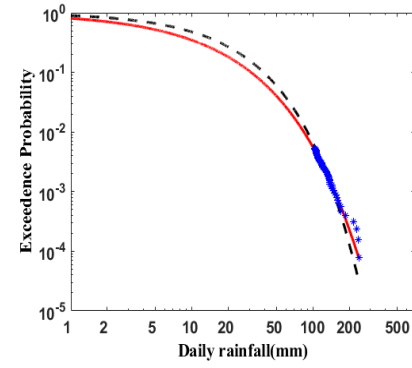
(iii) Weibull distribution (Station code: 497)
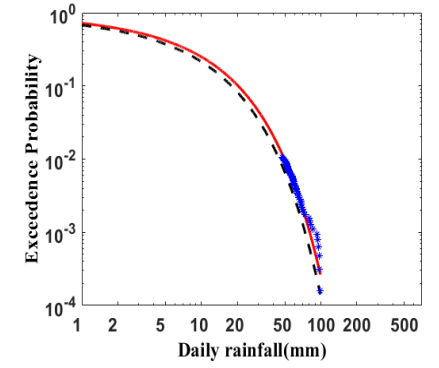
(iv) Gamma distribution (Station code: 525)

(i)
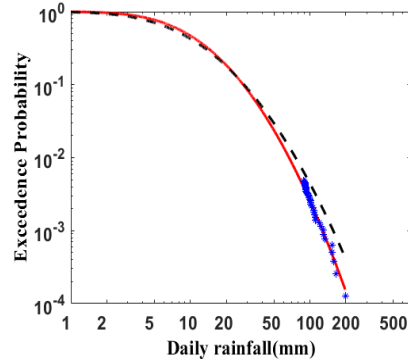
(i) Lognormal distribution (Station code: 26)
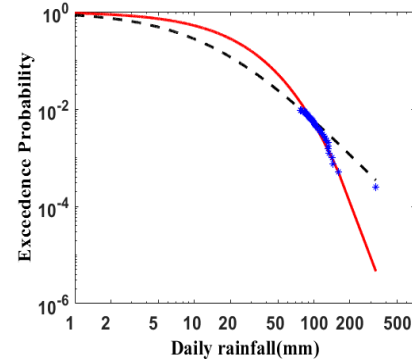
(ii) Pareto distribution (Station code: 14)
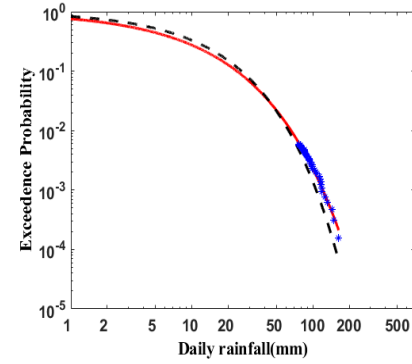
(iii) Weibull distribution (Station code:497)
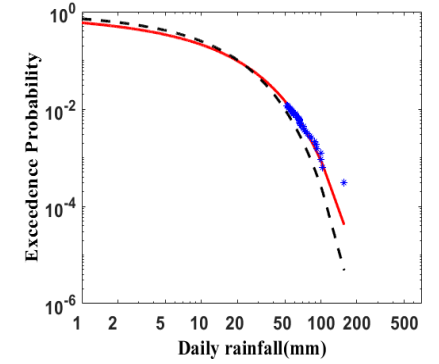
(iv) Gamma distribution (Station code: 525)

(ii)

Figure 3.2 PRMSE norm based fitting approach applied to four tails, namely Lognormal, Pareto II, Weibull, and Gamma for two different periods (i) pre-1970 (1901-1970), (ii) post-1970 (1971-2010)

3.3   Results

This chapter investigates the temporal and spatial changes in the behavior of daily extreme precipitation over India in terms of its frequency of occurrence due to the shift in the global climatic regime in the 1970s. The temporal changes are assessed between the two time periods, pre-1970 and post-1970, both at grid-scale and regional scale. Further, a categorical classification of grids based on the change in average rainfall above the threshold (increase or decrease in magnitude) and change in the nature of the tails (i.e., from light to heavy or heavy to light) over the two-time period is also presented in this study.

3.3.1   Assessment of spatial and temporal changes in tail behavior of probability distribution of daily precipitation at Grid-scale

In this section, the spatial and temporal changes in the behavior of tails of probability distributions of daily precipitation over India were analyzed for pre-and post-1970 periods. Following the procedure described earlier, Lognormal, Pareto type II, Weibull, and Gamma distributions were considered to fit the non-zero daily precipitation data at 4789 grids over India from 1901 to 1970 and 1971 to 2010, respectively. The distributions were fitted either by considering entire precipitation data at a grid or considering either 70 or 40 largest precipitation data values depicting the tail part of the distribution for the time periods of 1901 to 1970 and 1971 to 2010. Visual investigation of the fits at all 4789 grids indicated that the fit based on the largest values in the precipitation data adequately described the tail part of the empirical distribution. This shows the advantage of the threshold-based approach proposed by Papalexiou et al. (2013) for fitting probability distributions to daily precipitation data, especially while analyzing daily extreme precipitation events. To find the best-fitted distribution of the four fitted distributions at each grid, the PRMSE norm was considered in this study. In the case of each grid, the distribution function yielding the least estimate for the PRMSE norm was declared to be the best-suited distribution for that grid. Figure 3.3 shows the geographical or spatial variation of best-suited distribution over India for both pre-and post-1970 periods. For the pre-1970 period, out of 4789 grids over India, Lognormal distribution was found to be better suited for nearly 41.87% grids, followed by Pareto (32.43%), Weibull (18.56%), and Gamma (7.14%) distributions.
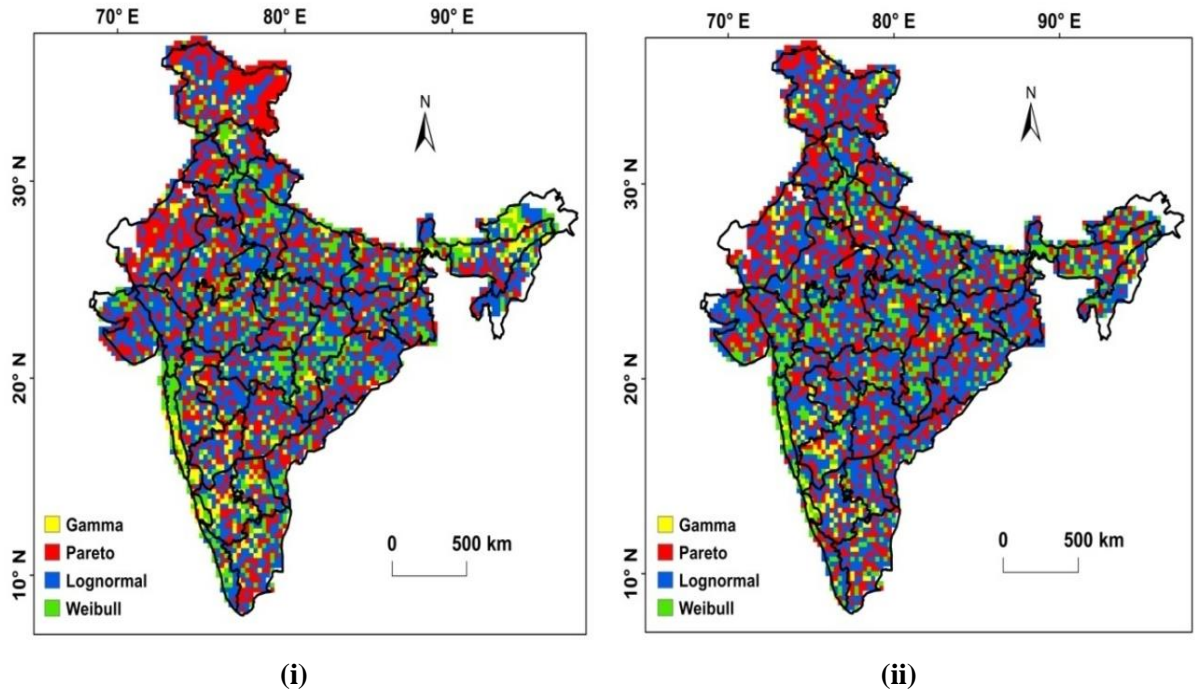
Figure 3.3 Geographical locations of best-fitted distribution tail over India for two time periods (i) pre-1970 (1907-1970), (ii) post-1970 (1971-2010). Different color coding has been used for different tail type

For the post-1970s, the sequence remains the same with Lognormal as the best-fitted distribution for most grids over India. It has been observed that 45.86% of grids exhibit lognormal distribution as best best-suited distribution, followed by Pareto for 32.20%, Weibull for 17.06%, and Gamma for 4.88% grids. Given these overall percentages, one may conclude that the Lognormal and Pareto type II distributions (both heavy-tailed distributions) are the most suitable distributions for modeling the tails of probability distributions of daily precipitation data. Overall, it can be seen that there is a 4% increment and 2.26% decrement in the number of grids following the Lognormal distribution and Gamma distribution, respectively. The Weibull distribution with a shape parameter less than 1 (which is a heavy-tailed distribution) was found to be suitable for 96.40% and 98.53% of the grids, which showed Weibull as the best-suited distribution for pre-and post-1970 periods, respectively. It can be concluded from the results that there exists a dominance of heavy-tailed distributions over light-tailed Gamma and Weibull distributions along with the increase in the tail heaviness of precipitation data over India due to climatic shifts. This indicates that extreme precipitation events in India have become more frequent. Details on the shape (α) and scale (β) parameters of the best-suited distributions are provided in Table 3.2 (i), (ii) for both pre- and post-1970 periods, respectively.

Table 3.2 Statistical summary based on fitting of the four distributions to the tails of precipitation data for (i) pre-1970, and (ii) post-1970 time periods

(i)

	Pareto			Lognormal		
	MSE	β	α	MSE	β	α
Min	0.0025	0.8866	0.0000	0.0026	1.5062	0.1878
Mean	0.1014	8.1352	0.2513	0.0581	10.2748	1.2208
Max	0.6572	25.6076	0.6910	0.8410	24.5317	2.1150
Median	0.0438	7.6815	0.2340	0.0335	10.0345	1.1873
SD	0.1224	4.2759	0.1206	0.0718	4.7997	0.2572
Skew	1.8099	0.5574	0.4701	3.4391	0.2278	0.4291
	Weibull			Gamma		
	MSE	β	α	MSE	β	α
Min	0.0041	0.3476	0.3087	0.0037	3.4304	0.0379
Mean	0.1108	12.2075	0.9143	0.1339	17.3272	1.0737
Max	0.9855	31.5581	17.7532	0.4853	30.1988	7.0028
Median	0.0737	12.4927	0.8501	0.1199	17.6201	0.9060
SD	0.1221	5.0817	0.7085	0.0826	4.2256	0.7182
Skew	3.4180	-0.1809	14.4425	0.8034	-0.2213	1.9645

(ii)

	Pareto			Lognormal		
	MSE	β	α	MSE	β	α
Min	0.0033	0.7920	0.0033	0.0040	1.9162	0.6009
Mean	0.0892	8.1501	0.2842	0.0551	9.6792	1.3141
Max	0.5203	31.8910	0.7445	0.4819	24.0319	2.2519
Median	0.0487	7.5730	0.2660	0.0377	9.1881	1.2814
SD	0.0967	4.2906	0.1314	0.0535	4.6877	0.2856
Skew	1.8672	0.6598	0.4807	2.6614	0.3851	0.4276
	Weibull			Gamma		
	MSE	β	α	MSE	β	α
Min	0.0030	0.3850	0.3066	0.0040	5.1253	0.1139
Mean	0.1173	12.2138	0.8655	0.1788	18.1197	1.1743
Max	0.9744	30.8359	19.2964	0.6899	32.7686	10.6443
Median	0.0862	12.6215	0.8158	0.1684	18.3819	0.9731
SD	0.1111	5.1673	0.6941	0.1021	4.3883	0.8208
Skew	3.3131	-0.1866	16.2781	0.6251	-0.1592	2.6873

The shape parameter is a scalar measure of tail behavior, and its histogram constructed based on estimates at all grids can be helpful in providing essential information about tail heaviness.

(i)   **(ii)**
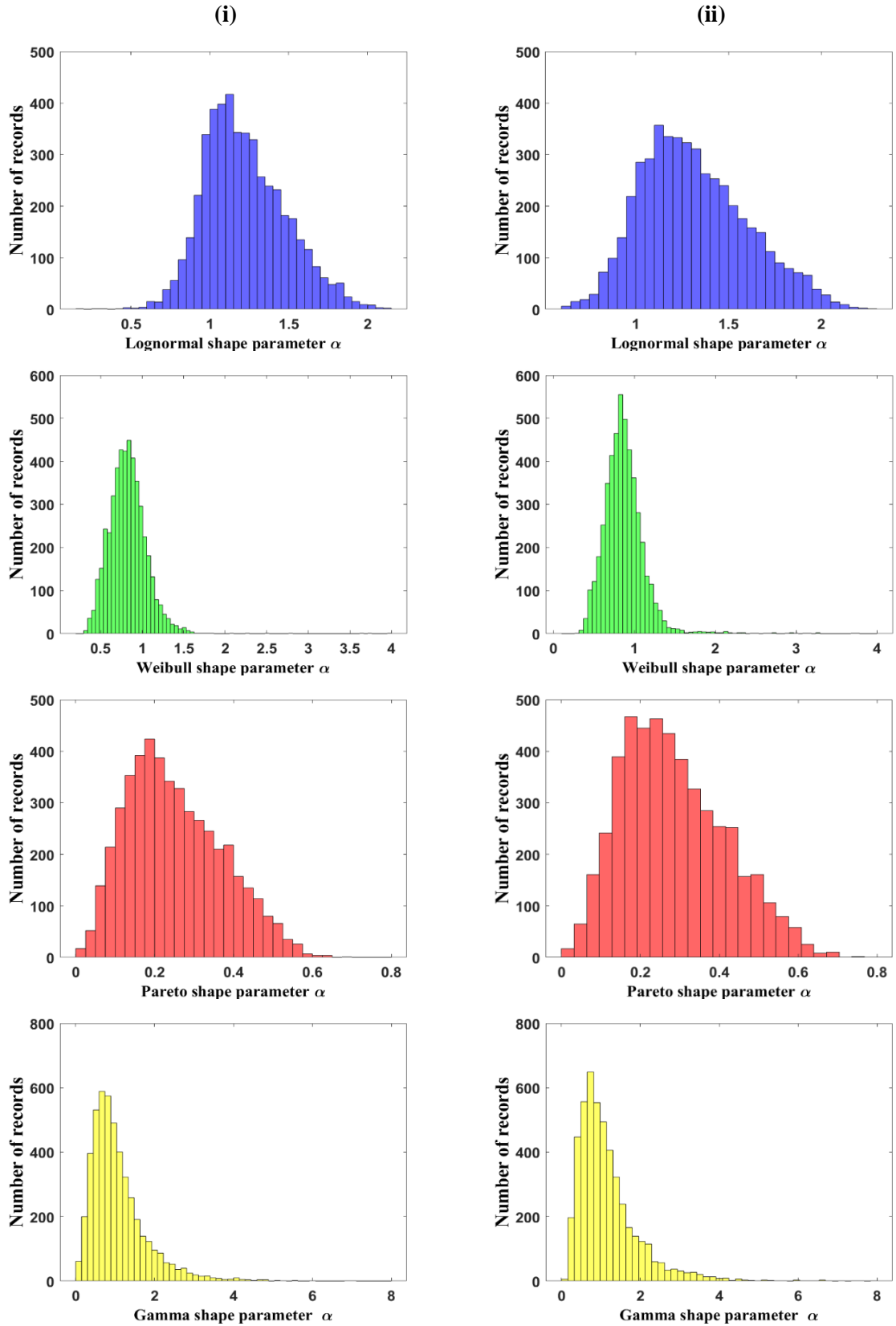
Figure 3.4 Histograms of the shape parameters of 4 distributions fitted to all 4789 records over two time periods (i) pre-1970 (1901-1970), (ii) post-1970 (1971-2010)

Figure 3.4 (i), (ii) shows the empirical histograms of the shape parameters of 4 distributions considered in this study for pre and post-1970s periods. Modal values of the histograms represent the most probable values of the shape parameters for each of the distributions. For Pareto type-II distribution, the modes were observed as 0.19 and 0.176 for the pre- and post-1970 periods. Low modal values for Pareto distribution imply the non-existence of statistical moments for higher orders, i.e., greater than 5.26 and 5.88 (Papalexious et al., 2013). The mode value of the shape parameter for Lognormal distribution was about 1.1 for both the pre- and post-1970 periods. In the case of Weibull distribution, the modes of the histograms were observed to be around 0.84 for pre-1970 and 0.82 for post-1970, both implying the presence of heavier tails of the distribution as the shape parameter is less than 1. Histograms for the shape parameter of Gamma distribution show low modal values of 0.67, and 0.73 for pre and post-1970s, respectively, which indicate the presence of hyper-exponential tails representing a lesser frequency of occurrence for extreme precipitation events.

Histograms of shape parameters did reveal a lot about the basic nature of the tail of four distributions but to further investigate the tail relevances in describing daily precipitation; the average ranking was also considered. All 4 distributions were ranked in the ascending order of the PRMSE norm, i.e., the distribution yielding the least PRMSE was declared as Rank 1 distribution while the distribution with the highest PRMSE was ranked as 4. Figure 3.5 illustrates the average rank of the four probability distributions for pre-and-post-1970 periods. A lower average rank of a probability distribution indicates better suitability of the distribution in describing the tails of precipitation data as compared to those with higher ranks. Lognormal distribution was the best-fitted distribution with an average rank of 1.9 and 1.7 for both pre-1970 and post-1970 periods. The best-fitted distributions were ordered as Lognormal, Pareto type II, Weibull, and Gamma distribution based on their ranks for both pre-and post-1970 periods. Conventionally, Gamma distribution is the most commonly used probability distribution for representing daily precipitation. However, results from this study inferred that the Gamma distribution was the worst performer for both periods. It can be noted from Figure 3.5 that the average ranking of Lognormal and Pareto type II distributions is decreased from 1.9 to 1.78 and 2.31 to 2.2 over the pre- and post-1970 periods, respectively. This shows that the gridded daily precipitation for the post-1970 period over India exhibits heavier tails than the pre-1970 period. Additionally, an increase in the tail heaviness of the distributions in the post-1970 period was also evident from the increase in the average rank of the Gamma distribution.

(i)

(ii)
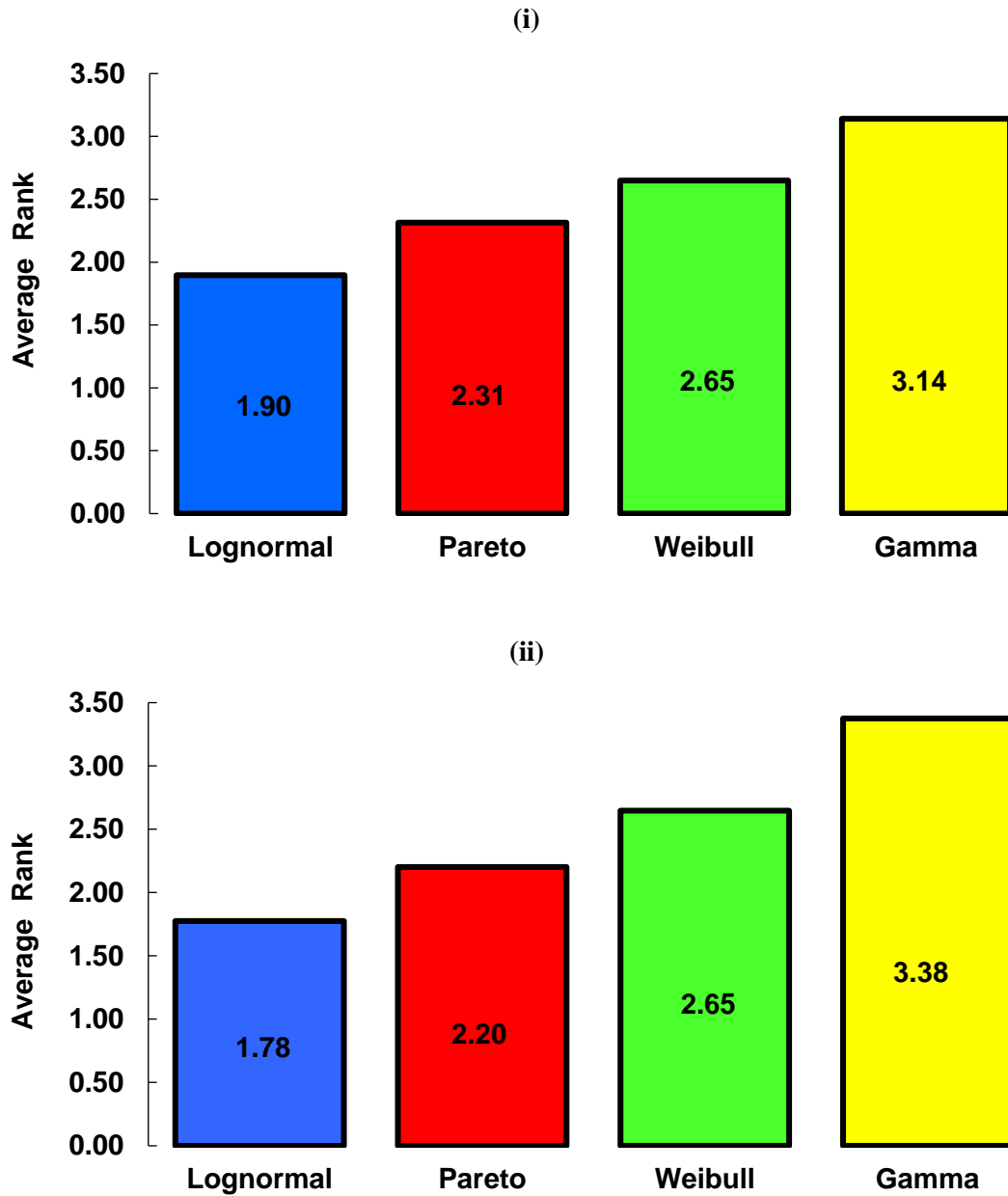
Figure 3.5 Mean ranks of 4 distributions tails for different time period (i) pre-1970 (1901-1970), (ii) post-1970 (1971-2010)

Another way adopted for assessing the temporal change in the tails of probability distributions during the pre-and post-1970 periods is achieved by comparing the tails of the distributions in couples or pairs. Various pairs of distributions considered in this study are "Lognormal *vs*. Pareto", "Pareto *vs*. Weibull", "Pareto *vs*. Gamma", "Lognormal *vs*. Weibull", "Lognormal *vs*. Gamma" and "Weibull *vs*. Gamma". The best-fitted distribution among the pair (any two distributions) was selected based on the PRMSE norm for each grid. The distribution with a lesser PRMSE value was considered as the best fit. Figure 3.6 illustrates the comparison between 2 probability distributions in couple for pre-and post-1970 periods. The figure presents the percentage of grids found suitable for each probability distribution compared in pairs. It can be deduced from the figure that the Lognormal distribution (which is a heavy-

tailed distribution) fits the extreme daily precipitation data for 58.66% and 60.51% grids in the pre-and post-1970 periods when compared with the Pareto type II distribution.
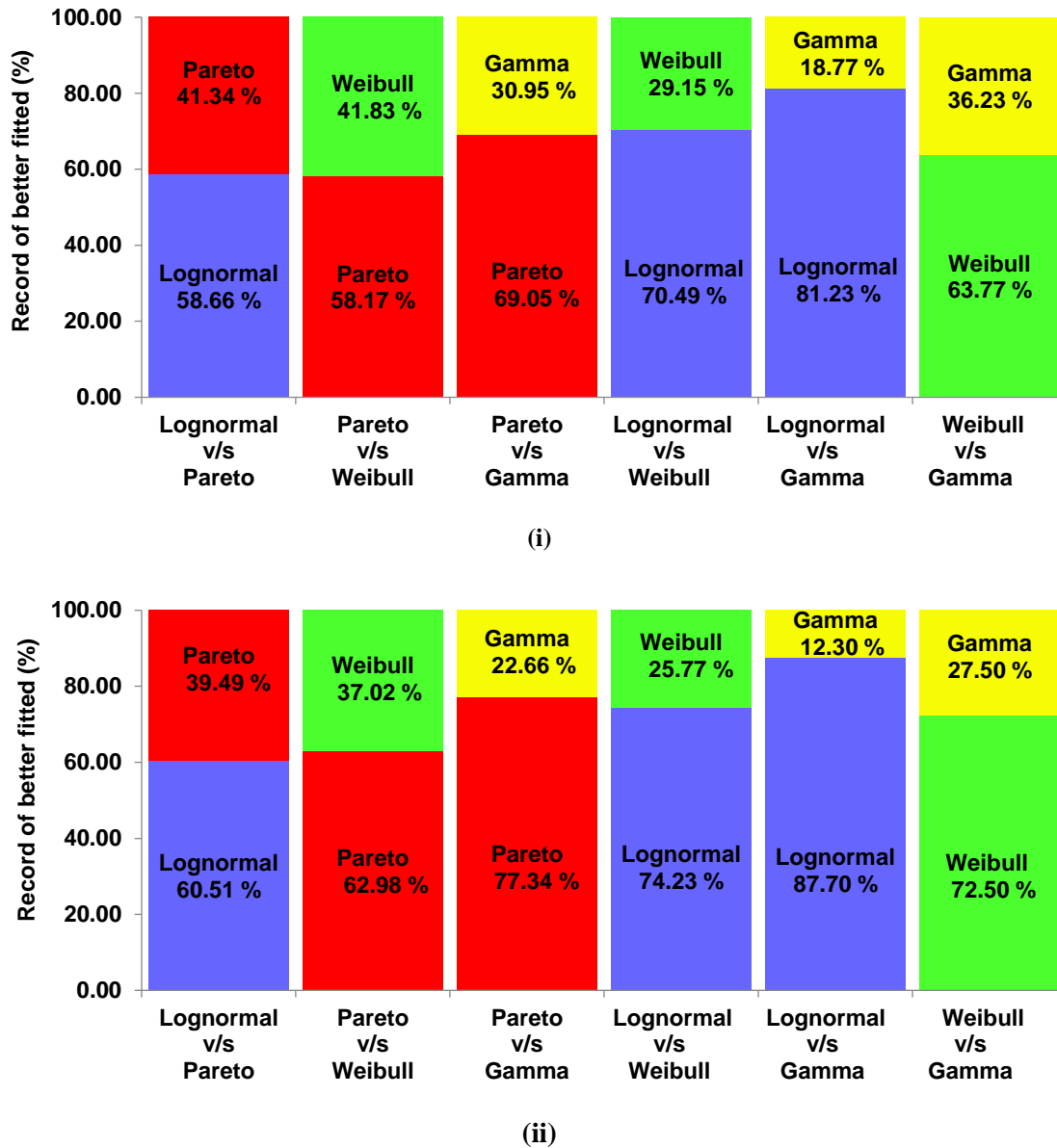
Figure 3.6 PRMSE norm-based comparison of the fitted tails in couples for two time period (i) pre-1970 (1901-1970), (ii) post-1970 (1971-2010)

Further, the Lognormal distribution was found to be better suited than the Weibull and Gamma distributions for both periods. It can be noted that the percentage of grids where daily precipitation is well represented by the Lognormal distribution against the Weibull and Gamma distributions has increased from 70.49% to 74.23% and 81.23% to 87.7%, respectively, over the pre-and post-1970 periods. This indicates that the probability distributions of daily precipitation in the post-1970 period exhibit heavier tails than the pre-1970 period. We have also compared the Pareto type II distribution with Lognormal, Gamma, and Weibull distribution, and it emerged as the second best-fitted distribution

after Lognormal distribution. Similar to Lognormal distribution, Pareto type II distribution was inferred as a better fitting distribution against Gamma and Weibull distributions for both pre-and post-1970 periods. The analysis revealed that the percentage of grids where the tails of probability distributions of daily precipitation are better fitted by the Pareto type II distribution against the Weibull and Gamma distributions has increased from 58.17% to 62.98 % and 69.05% to 77.34% over the periods. Amongst the Weibull and Gamma distributions, the Weibull distribution was better suited for describing the tails of daily precipitation data over India in both pre-and post-1970 periods. Interestingly, a heavier-tailed distribution was better fitted in each case during both periods. These findings point out that the heavier-tailed distributions should be preferred over their counterparts while representing the tails of daily precipitation data over India.

We have also investigated the existence of any geographical/spatial pattern of best-suited distributions over India. The maps shown in Figure 3.3 illustrate the spatial distribution of best-fitted distributions for pre-and post-1970 periods. These maps do not unveil any regular patterns; instead, they seem to follow a random spatial variation. Hence, to reveal some meaningful conclusions, we categorized the best-suited distributions into either sub-exponential or exponential-hyperexponential classes based on the estimates of the shape parameter (α), following El Adlouni et al. (2008) and Papalexious et al. (2013). The subexponential class includes Pareto type II distribution, Lognormal distribution, and Weibull distribution with shape parameter $\alpha < 1$, while the exponential-hyperexponential class comprises the Gamma distribution and Weibull distribution with shape parameter $\alpha > 1$. Figure 3.7 represents maps showing the spatial distribution of sub-exponential and exponential-hyper-exponential distributions over India for pre-and post-1970 periods. For the pre-1970 period, subexponential distributions were better suited for 4415 grids out of 4789 grids (i.e., about 92.19%) over India. In comparison, the latter class was found to be applicable for merely 374 grids (i.e., only 7.81% of all the grids). Similarly, for the post-1970 period, subexponential distributions were found to be adequate to model daily precipitation data at 4543 grids (i.e., 94.86% of all the grids while exponential-hyperexponential tails were found to be appropriate for the remaining 246 grids (i.e., 5.14% of the grids). It can be observed from maps that the heaviness in the tails of probability distributions of daily precipitation over India has increased post-1970s climatic shift. With the dominance of heavy tails all over India, few pockets comprising lighter tails were observed in the northeast region and along the western coastal plain of the Indian Peninsular for both periods. It can very well be seen from the maps that the subexponential tails are much more dominant over Indian regions than the exponential-hyperexponential tails.

Overall, the comparison is made in terms of the difference in the percentage of the number of grids belonging to one category and the values of the shape parameter (α). The maps showing the distribution and the two broad classes of tail behavior (i.e., sub-exponential and exponential-hyperexponential) are compared for both pre-and post-1970 periods to examine the impact of the global

shift in climate regime in the 1970s on extreme precipitation. The presence of heavy tails in the daily precipitation data points to the fact that extreme precipitation events over India are no longer rare.
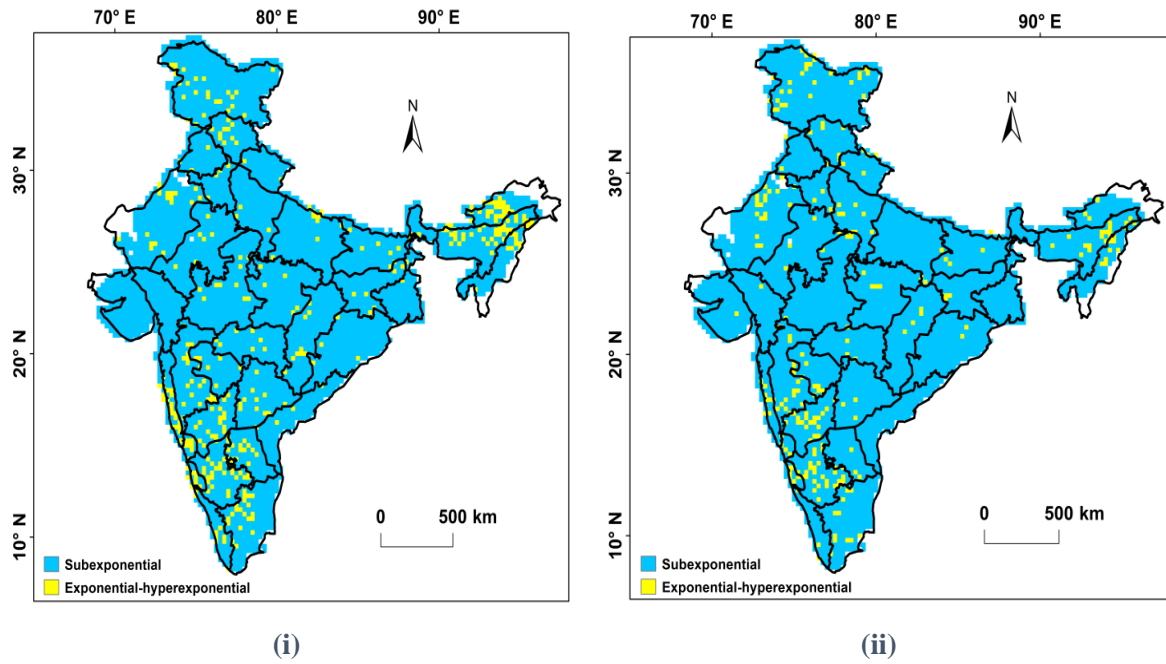
Figure 3.7 Geographical Variation of Subexponential and Exponential-hyperexponential tails over India for different time period (i) pre-1970 (1901-1970), (ii) post-1970 (1970-2010)

3.3.2 *Assessment of temporal changes in tail behavior of probability distribution of daily precipitation at Regional Scale*

Analysis at the grid scale revealed some essential inferences about the tail behavior of daily precipitation over India. However, to make the analysis more interpretable and usable at the regional scale, the temporal changes in tail behavior of the probability distribution of daily precipitation was assessed over 34 Meteorological Subdivisions over India. Table 3.3 provides details about the percentage of grids having heavy or subexponential tails in each subdivision, considering the pre-1970 and post-1970 periods. Figure 3.8 shows the percentage of subexponential tails in each Meteorological Subdivision over India for both periods using color codes. In the case of the pre-1970 period, Saurashtra Kutch and Diu (subdivision 22) showed a complete dominance of heavy tails, followed by Gujarat (subdivision 21), Gangetic West Bengal (subdivision 6), and Orissa (subdivision 7).

Table 3.3 Summary of percentage of the grids having heavy tails within each Meteorological Subdivision for the pre and post-1970 records

ID	Metrological Region	Percentage of grids having heavy tails	
		Pre-1970 (1901-1970)	Post-1970 (1971-2010)
2	Arunachal Pradesh	68.54	95.51
3	Assam and Meghalaya	77.71	90.36
4	Naga Mani Mizo and Tripura	88.89	90.00
5	Sub Him W Bengal Sikkim	96.67	100.00
6	Gangetic West Bengal	98.92	100.00
7	Orissa	98.15	99.07
8	Jharkhand	93.86	97.37
9	Bihar	94.90	96.82
10	East Uttar Pradesh	95.48	99.10
11	West Uttar Pradesh	98.05	94.81
12	Uttaranchal	97.59	93.98
13	Haryana Chandigarh and Delhi	93.20	94.20
14	Punjab	96.67	97.78
15	Himachal Pradesh	87.64	93.18
16	Jammu and Kashmir	93.83	93.09
17	West Rajasthan	93.18	92.80
18	East Rajasthan	97.09	98.06
19	West Madhya Pradesh	95.95	98.38
20	East Madhya Pradesh	97.45	97.45
21	Gujarat	99.19	97.58
22	Saurashtra Kutch and Diu	100.00	99.36
23	Konkan and Goa	71.43	90.00
24	Madhya Maharastra	87.50	88.82
25	Marathwada	93.26	92.13
26	Vidarbha	94.89	97.08
27	Chhatisgarh	95.00	97.78
28	Coastal Andhra Pradesh	96.80	100.00
29	Telangana	92.67	98.00
30	Rayalaseema	84.44	95.56
31	Tamil Nadu and Pondicherry	88.57	93.71
32	Coastal Karnataka	58.97	74.36
33	North Interior Karnataka	82.69	81.73
34	South Interior Karnataka	78.05	78.05
35	Kerela	86.15	95.38

Nearly 22 subdivisions had more than 90% of grids showing heavy-tailed behavior. For the post-1970 period, Sub Him West Bengal and Sikkim (subdivision 5), Coastal Andhra Pradesh (subdivision 28), and Gangetic West Bengal(subdivision 6) have 100% heavy tails. In the post-1970s, 30 subdivisions were found to have more than 90% of grids showing heavy tail behavior.
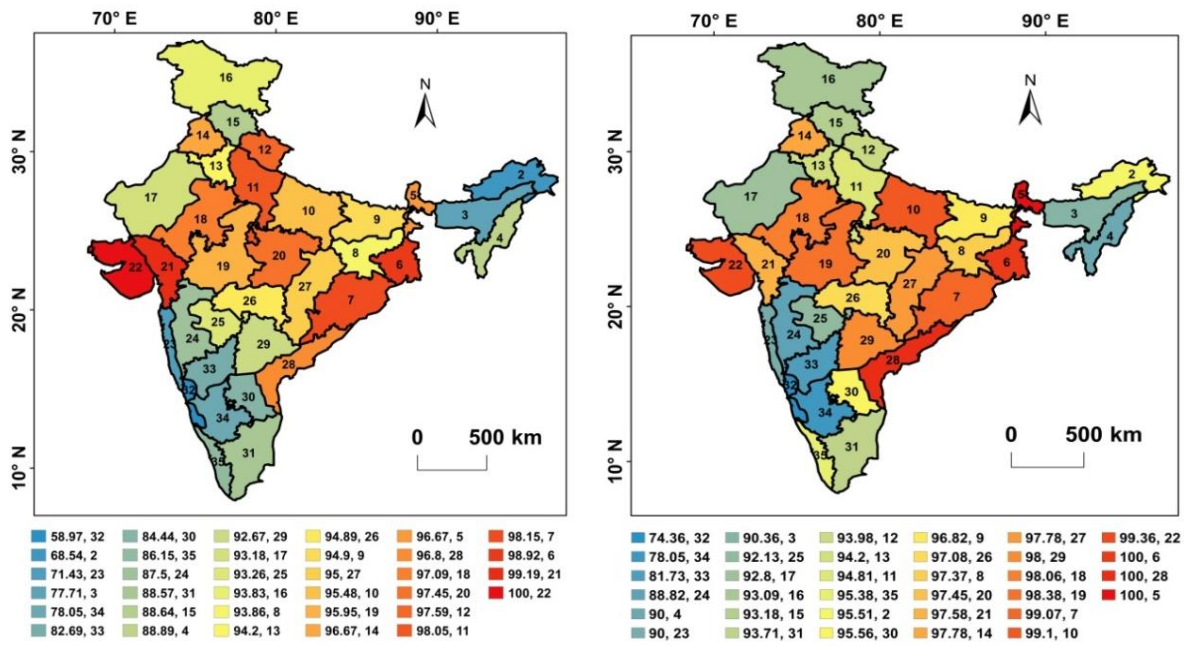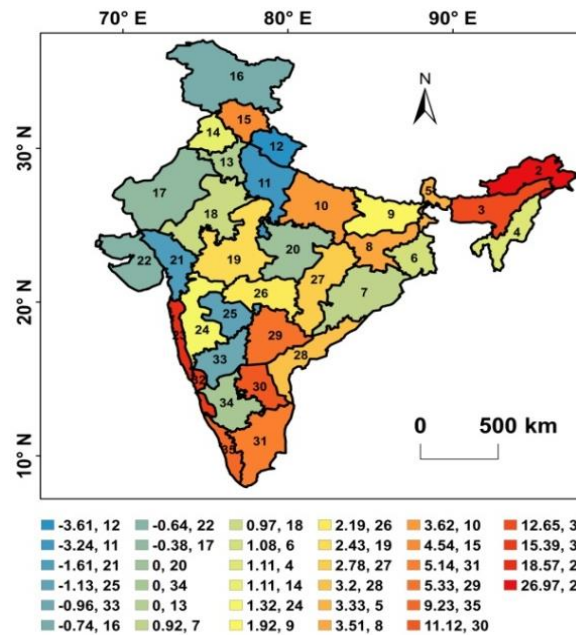
Figure 3.8 Geographical variation of percentage of subexponential tails in each Meteorological Subdivision over India for a different period (i) Pre-1970 (1901-1970), (ii) Post-1970 (1970-2010). Further, changes in the percentage of grids showing heavy tails due to climate shift in the 1970s are presented in subfigure (iii)

The maps showing the percentage of heavy tails in each subdivision for pre- and post-1970 were compared to find the change in the number of grids comprising heavy tails. The changes in the percentage of heavy tails over time for each subdivision are presented in Figure 3.8 (iii). In nearly 23

out of 34 subdivisions, the percentage of grids with heavy tails is observed to increase over time. Out of those 23, 9 subdivisions, namely Arunachal Pradesh, Assam and Meghalaya, Himachal Pradesh, Konkan and Goa, Telangana, Tamil Nadu and Pondicherry, Rayalaseema, Coastal Karnataka, and Kerela, showed an increase of about 5% or above. The substantially higher percentage of grids exhibiting heavy tails post-1970 compared to that of pre-1970 might be a possible consequence of climate change and global climatic shifts in the 1970s.

3.3.3 Assessment of Temporal changes in Magnitude and Frequency of Extreme Precipitation over India

The classification of grids exhibiting the severity in terms of increase in the magnitude and frequency of extreme precipitation events due to the climatic shift is achieved by considering the combined effect of change in average precipitation above threshold and change in the tail behavior. The rainfall values above the threshold are the ones that belong to the tail. We considered the average of these values at each grid, which served as an indicator of the magnitude of the extreme precipitation. The average rainfall values above the threshold vary from a minimum value of 25.19 mm to 287.51 mm for the pre-1970 period. On the other hand, average rainfall values above the threshold range from 39.65 mm to 646.71 mm.

Table 3.4 Eight categories of severity proposed by considering an increase or decrease of average rainfall above threshold and change in the nature of tail over the period from pre-1970 to post-1970

Category (Severity decreases top to bottom)	Change in average rainfall above the threshold from pre-1970 to post-1970	Change in tail type from pre-1970 to post-1970	Number of grids (out of 4789)	Percentage of grids (%)
Category 1	Increases	Heavy to heavy	3093	64.58
Category 2	Increases	Light to heavy	260	5.42
Category 3	Decreases	Heavy to heavy	1118	23.34
Category 4	Decreases	Light to heavy	72	1.50
Category 5	Increases	Heavy to light	98	2.04
Category 6	Decreases	Heavy to light	106	2.21
Category 7	Increases	Light to light	28	0.58
Category 8	Decreases	Light to light	14	0.29
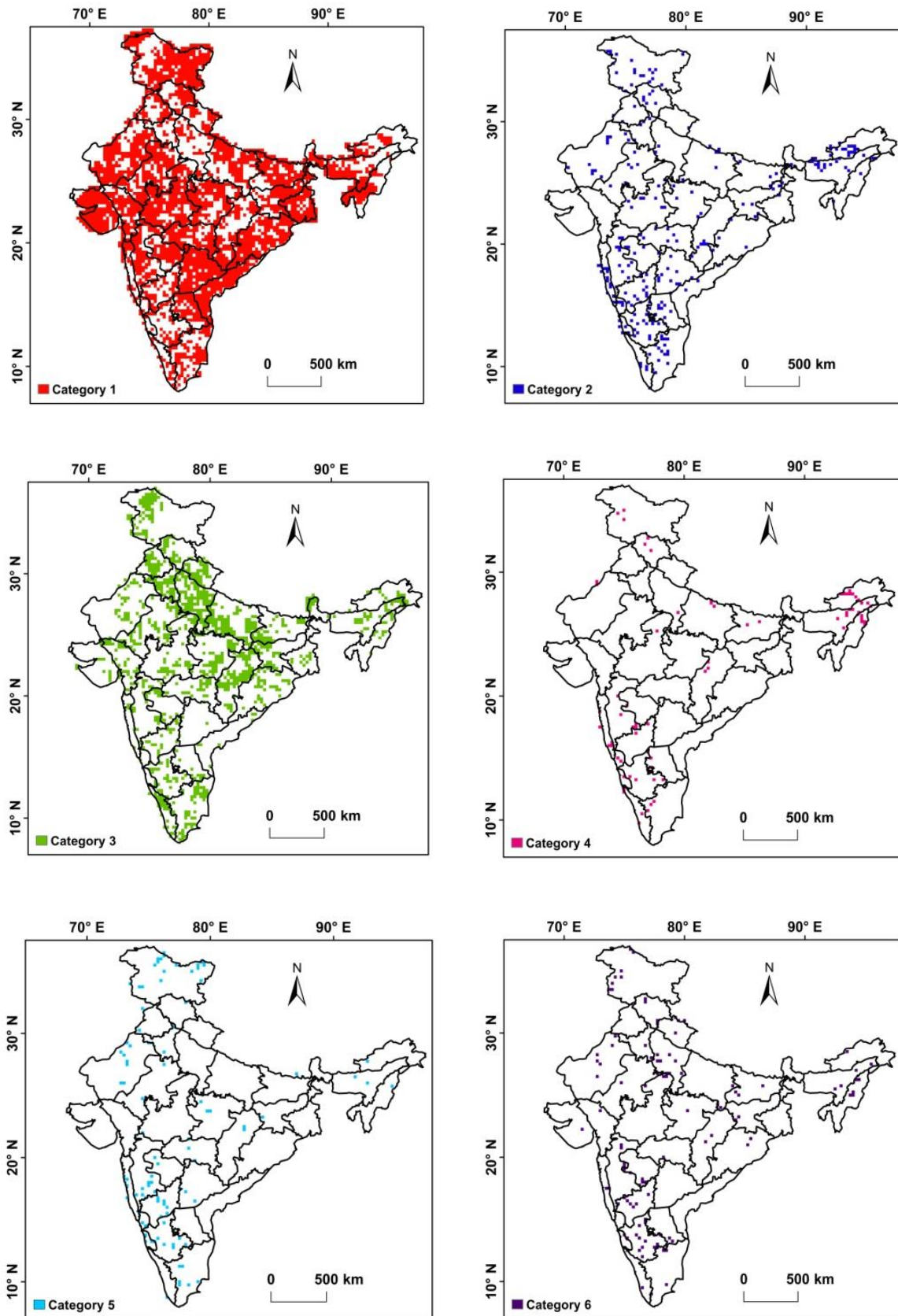
Figure 3.9 Spatial pattern of grids belonging to 8 Categories describing severity in terms of change in increase or decrease of average extreme rainfall above threshold and change in the nature of tail over the period from pre-1970 to post-1970
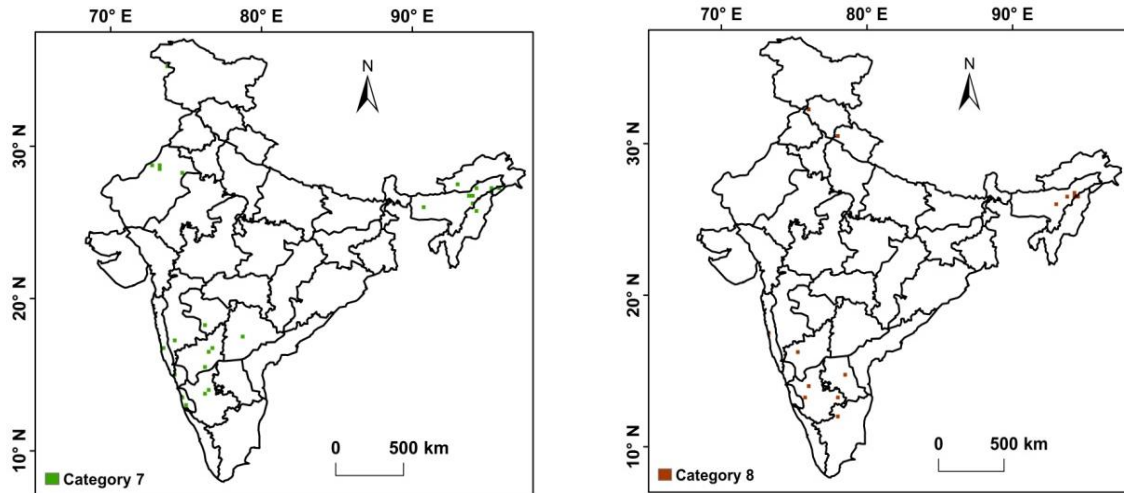
Eight categories of severity were proposed by considering an increase or decrease of average rainfall above the threshold and a change in the nature of the tail over the period from pre-1970 to post-1970. Table 3.4 describes the categories along with the number and percentage of grids falling in them. The categories are ranked from 1 to 8, with 1 being the most severe case and 8 being the least severe case. Figure 3.9 shows the spatial pattern of grids belonging to each category. Figure 3.9 and Table 3.4 show that nearly 64.5% of grids belong to Category 1, representing the most severe case. Category 2 comprises 5.42% of grids where an increase in the magnitude of extreme precipitation and transition in the tail behavior from light to heavy were observed. Category 3 includes nearly 23.34% grids where the magnitude of extreme precipitation has decreased with heavy-tailed behavior during pre- and post-1970. Overall, most of the grids in India show a tendency to transition towards heavier tails along with an increase in the magnitude of extreme precipitation.

3.4   Summary and conclusions

In this chapter, we have analyzed the temporal and spatial change in the tail behavior of daily precipitation over India from pre- to post-1970 time periods as per the global climatic shift. The tail behavior of precipitation data is assessed by identifying the best-fitted distribution out of 4 theoretical distributions to the sample obtained by the Annual Exceedance Series (AES) approach (e.g., Pareto type II, Lognormal, Weibull, and Gamma distributions) based on PRMSE norm. The approach is found to be easy to use and effective in diagnosing the tail behavior of daily precipitation data. Maps showing the geographical variation of the percentage of best-fitted sub-exponential tails over 34 Meteorological Subdivisions in India are given in this study. Also, the categorical classification of grids in terms of severity by considering the combined effect of an increase or decrease in average rainfall above the threshold and change in the nature of the tail over the period from pre-1970 to post-1970. Results from this study emphasize the importance of heavy-tailed distributions for reliable estimation of the frequency of extreme precipitation events in India. Important highlights from this study are as follows.

(i)   Lognormal and Pareto type II distributions (both heavy-tailed distributions) are found to be better suited for daily precipitation over India for both pre-and post-1970 periods. It can be concluded from the results that there exists a dominance of heavy-tailed distributions over light-tailed Gamma and Weibull distributions along with the increase in the tail heaviness of precipitation data over India due to climatic shifts. This directs to the fact that extreme precipitation events in India have become more frequent during both the pre-and post-1970 periods.

(ii)   Gamma distribution, in general, underestimates the frequency and magnitude of extreme events. Hence the distribution should not be considered for modeling the extreme precipitation events over India.

(iii)   Histograms of shape parameters of the 4 probability distributions revealed that the tails of daily precipitation data have become heavier from pre- to post-1970 periods.

(iv)   Heavy-tailed distributions can describe the observed precipitation extremes more effectively than light-tailed distributions. About 92.19% of the records in the pre-1970s and 94.86% in the post-1970s are better characterized by subexponential tails. Exponential-hyperexponential tails are found to be better suited for only 7.81% and 5.14% records for the pre-and-post-1970s periods. It can be seen that increasing trends of heavy tails persist in the later period indicating a rising trend of more frequent and 'severe events' of precipitation.

(v)   Twenty-three meteorological subdivisions in India show an increase in the percentage of heavy tails in the post-1970s compared to the pre-1970s. Further, 9 subdivisions out of those 23, namely Arunachal Pradesh, Assam and Meghalaya, Himachal Pradesh, Konkan and Goa, Telangana, Tamil Nadu and Pondicherry, Rayalaseema, Coastal Karnataka, and Kerela, showed a substantial increase in the percentage of grids exhibiting heavy tails.

(vi)   Eight categories of severity are proposed by considering an increase or decrease in average rainfall above the threshold and change in the nature of the tail over the period from pre-1970 to post-1970. Nearly 70% of grids in India belong to Category 1 and Category 2 which are deemed to indicate severe/critical categories in terms of increase in the magnitude of extreme precipitation and the presence of heavier tails over pre- to post-1970 periods.

An important inference from this analysis is that the frequency and the magnitude of extreme precipitation events have generally been undervalued in the past. The use of light-tailed distributions for modeling daily precipitation can lead to a serious underestimation of the frequency and the magnitude of design extreme precipitation, which is highly undesirable for the design of water control structures. It can be noted that the results obtained from the present study are dependent on the length of the precipitation record (e.g., Arguez and Vose, 2011; Cavanaugh et al., 2015) available at each grid and the presence of serial dependence among the peaks/extreme precipitation events selected using the AES approach (e.g., Koutsoyiannis, 2008). Extended research is underway to alleviate the limitation of serial

dependence among the selected extreme precipitation events in the AES by exploring the strategies that can form a sample with independent events (e.g., Adams et al., 1986).

CHARACTERIZING THE TAIL BEHAVIOR OF DAILY PRECIPITATION PROBABILITY DISTRIBUTIONS OVER INDIA USING THE OBESITY INDEX

4.1 Overview

Extreme precipitation events lie in the upper part of the probability distribution of daily precipitation data, i.e., the tail. Prediction of extreme precipitation depends on how reliably the distribution tail is modeled. Tail behavior can be studied by graphical as well as threshold-based fitting approaches. However, the graphical methods are time-consuming and do not provide quantitative comparisons between two distributions, whereas the threshold-based approaches possess limitations such as ambiguity in selecting an optimum threshold for demarcation of the tail. There have been attempts in the past to define and use different upper tail indicators like Shape parameter, Gini Index, and Upper tail ratio (UTR) to characterize the heavy tail phenomenon. This chapter presents a novel index, known as "Obesity Index" (OB), which is a quantitative, scalar, and easy-to-use indicator that can diagnose the heaviness of distribution tails without assuming any threshold for segregating the tail. This indicator alleviates the limitations associated with the threshold-based and graphical approaches. This chapter presents the utility of the concept of OB for performing a diagnosis of the heaviness of tails and characterization of the daily gridded precipitation data over India for historical as well as future time period. A step-by-step algorithm for the determination of the OB using the bootstrapping technique for reliable discrimination of various distribution tails has been proposed. The robustness of the approach is established through a simulation study by investigating the adequate sample length and the optimum number of random samples required for the application of the algorithm. The OB approach as well as the threshold-based approach is applied to characterize the tail behavior of daily gridded precipitation over India. Subsequently, a comparative analysis between the OB-based approach and threshold-based approaches proposed by Papalexiou et al. (2013) and Nerantzaki and Papalexiou (2019) is performed to examine their ability to characterize the probability distributions of the daily gridded precipitation over India into two broad classes of distribution i.e., sub-exponential and exponential/hyper-exponential. The approach was also used to examine the change in OB over time by considering various climate change scenarios. Further, the application of the OB-based approach is extended to characterize daily precipitation in Indian Meteorological Subdivisions. Finally, the spatial variation of OB with respect to the elevation was explored and an attempt was made to relate it to the climatology of India.

4.2 Theoretical Background on Obesity Index

The concept of the OB was proposed based on the heuristics that a heavy-tailed sample usually has a few large values compared to the other values (Cooke and Nieboer, 2011; Cooke et al., 2014). The concept overcomes the limitations associated with graphical or threshold-based approaches (Wietzke et al., 2020). The major advantage associated with the OB is that it does not require assuming any threshold for segregating the tail part of a distribution (i.e., it can measure the heaviness of the distribution tail without referring to the limiting behavior of a distribution) (Nieboer, 2011; Cooke et al., 2014). This approach checks how far the largest sample values (here, daily non-zero precipitation data) occur from the main body of the probability distribution, also known as the principle of "a single big jump". This principle forms a basis for determining the presence of sub-exponentiality in the data. Further, the index is an empirical measure and can be computed directly from the data without assuming any form of distribution function. The limitation associated with the concept of OB is that it is applicable for a random variable that can take only positive values. Another limitation is that for a symmetric distribution like Normal (which is a light-tailed distribution) and Cauchy distribution (which has a heavy tail), OB takes the same value equal to 0.5.

Mathematically, the OB can be defined for a positive random variable X as,

$$OB(X) = P(X_1 + X_4 > X_2 + X_3 \mid X_1 \leq X_2 \leq X_3 \leq X_4) \tag{4.1}$$

Where, $\{X_1, \ldots, X_4\}$ are independent copies of X. When X_4 is very large compared to others X_S, then the probability that the sum of largest and smallest observation in a dataset of four is greater than the sum of the other two observations as represented by Equation 4.1. This probability tends to be higher for datasets having the presence of large values or extremes, especially in the case of heavy-tailed distributions. This property can form the basis for distinguishing between light and heavy-tailed distributions. Being an estimate of probability, the OB takes values from 0 to 1.

A step-by-step algorithm for determining the OB using the bootstrapping technique (Efron, 1979; Daníelsson et al., 2001; Qi, 2008) is proposed below.

1. Generate a sample of the random variable X having a sample size equal to N
2. Define Max_count as the maximum number of times for which random sampling can be performed
3. Initialize iteration count $= 1$
4. Resample $X_i^{count}s$ for $i = 1, \ldots, 4$ from the generated sample which represents independent copies of X
5. Sort the $X_i^{count}s$ in ascending order, i.e., $X_1^{count} \leq X_2^{count} \leq X_3^{count} \leq X_4^{count}$
6. Verify whether $X_1^{count} + X_4^{count} > X_2^{count} + X_3^{count}$. If $X_1^{count} + X_4^{count} > X_2^{count} + X_3^{count}$ then denote $K(count) = 1$ else 0

7. If, $count < Max\_count$ then increment $count = count + 1$ and repeat steps 4 to 6. If $count = Max\_count$, then proceed to step 8

8. Determine OB of the data using Equation (4.2)

$$\Rightarrow OB(X) = \frac{\sum_{count=1}^{Max\_count} K(count)}{Max\_count}$$

9. If, $OB(X) \le 0.75$, then the random variable X has a light-tailed distribution. Otherwise (i.e., $OB(X) > 0.75$), X is considered to be originated from a heavy-tailed distribution

Intuitively, the OB is an empirical estimate that can be linked to the tails of various distributions. The greater the value of OB, the heavier the tail of a distribution. Obesity Index behavior for various probability distribution functions is further investigated in sub-section 4.2.1 by analyzing its variation with respect to distribution shape and scale parameters.

4.2.1   Behavior of obesity index with shape parameter of probability distributions

In this section, the behavior of the OB for the variation in shape and scale parameters of various distributions is investigated. For this, a simulation study was performed by considering samples from 4 probability distribution functions viz., Pareto, Weibull, Lognormal, and Gamma following Papalexiou et al. (2013, 2018), Gupta et al. (2021). Details about the distributions can be found in Table 3.1 of Chapter 3. Samples of size 20,000 were generated by considering the four distribution functions. Subsequently, for each of the distributions, the estimates of the OB were computed following the steps mentioned in the algorithm by considering several combinations of shape and scale parameters. Graphs showing the variation of OB with respect to shape and scale parameters for the distributions are provided in (Figure 4.1). The graphs corresponding to Lognormal and Pareto distributions illustrate a concave increase, while those corresponding to Weibull and Gamma distributions show a convex decrease for the OB with respect to an increase in shape parameter. The equations of 3rd order polynomials were provided in the figure depicting the nonlinear relationship between the shape parameter and OB. However, it can be observed from the graphs that the change in the scale parameter of distribution does not have much influence on OB. Figure 4.1(ii) shows that the OB for a particular case of Gamma distribution with shape parameter as 1 (i.e., Exponential distribution) takes a value equal to 0.75. Therefore, when the obesity index takes a value greater (lesser) than 0.75, the samples can be assumed to come from a heavy (light)-tailed distribution. Papalexiou et al. (2013) stated that the Pareto type II and Lognormal distributions belong to the sub-exponential class and are considered as heavy-tailed distributions. Weibull can belong to both classes (i.e., light as well as heavy tail), depending upon the values of its shape parameter, while the Gamma distribution has essentially an exponential tail. One can infer from Figure 4.1 that except for the Pareto type II distribution, other distributions can possess heavier and lighter tails depending upon the values taken by the shape parameter. The obesity/heaviness

of the distribution tails can be quantified based on the value taken by the OB (refer to Figure 4.1). It can be observed from the figure that Weibull or Lognormal distributions can possess more obese or heavy tails than a Pareto distribution, depending upon the choice of parameters, especially the shape parameter.

Additionally, we investigated the sensitivity of precipitation quantiles corresponding to various non-exceedance probabilities ($F=$ 0.95, 0.97, 0.99, and 0.999) to the variation in the OB. For this analysis, we considered three heavy-tailed distributions, *viz.* Pareto type II, Lognormal, and Weibull distributions, and the scale parameter as 0.5, 10, and 25. Obesity Index is varied from 0.75 to 0.95 with an increment of 0.01 as this range generally belongs to heavy-tailed distributions. For each of the distributions, the estimate of the shape parameter corresponding to the value of OB was interpolated from the plots given in Figure 4.1 (corresponding to the scale parameter under consideration). Figure 4.2 shows the sensitivity of precipitation quantiles to the change in OB. Note that the quantiles shown on the *y*-axis are in a logarithmic scale. It can be observed from the figure that there is considerable variation in precipitation quantile estimates when the OB is even changed by 0.02. Precipitation quantiles increase exponentially when the OB estimate becomes greater than 0.85.
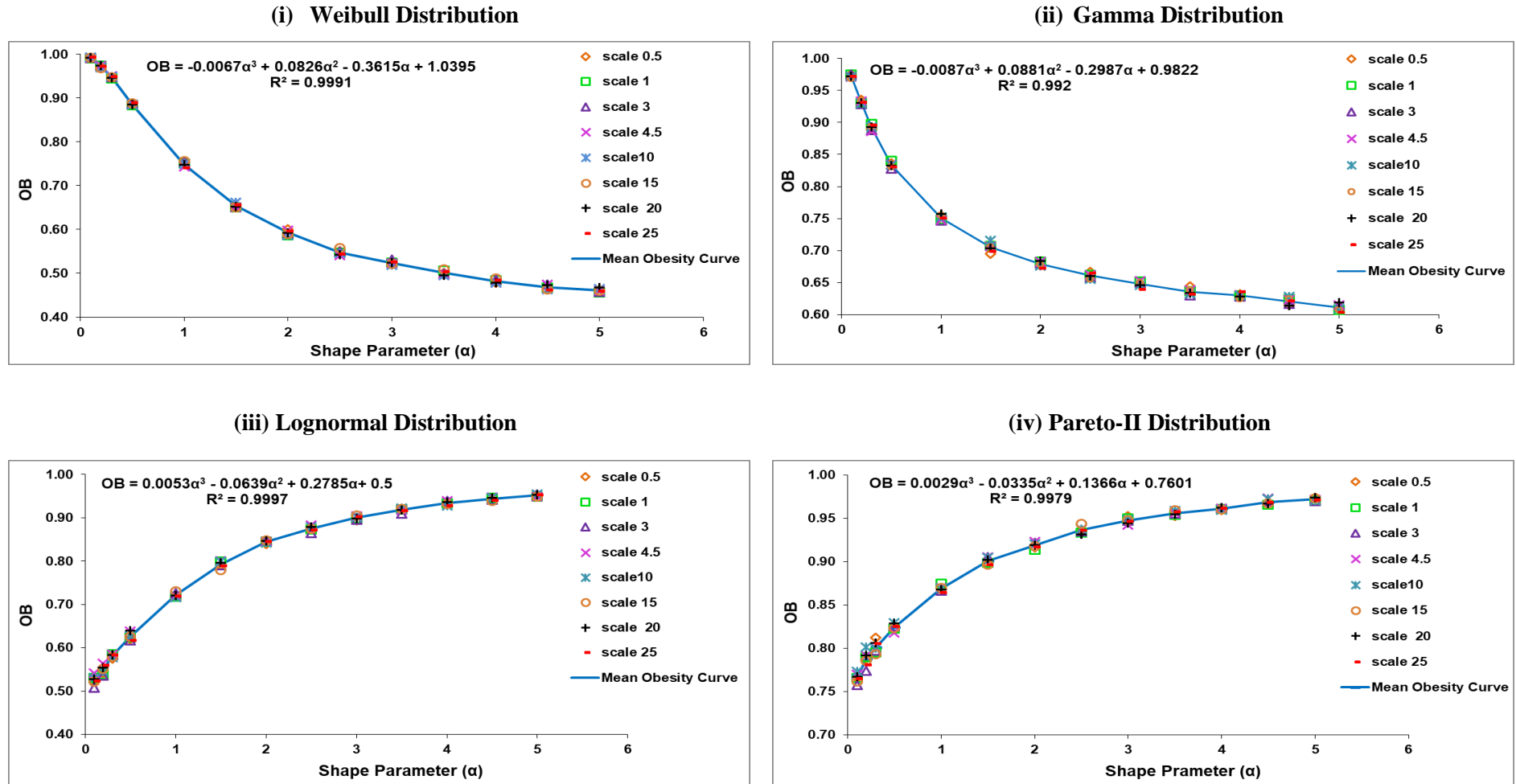
Figure 4.1 Variation of obesity index with respect to shape and scale parameters for (i) Weibull, (ii) Gamma, (iii) Log-Normal, and (iv) Pareto Type II distributions. The colored curves represent the variation of the mean of OB over shape parameter (α)
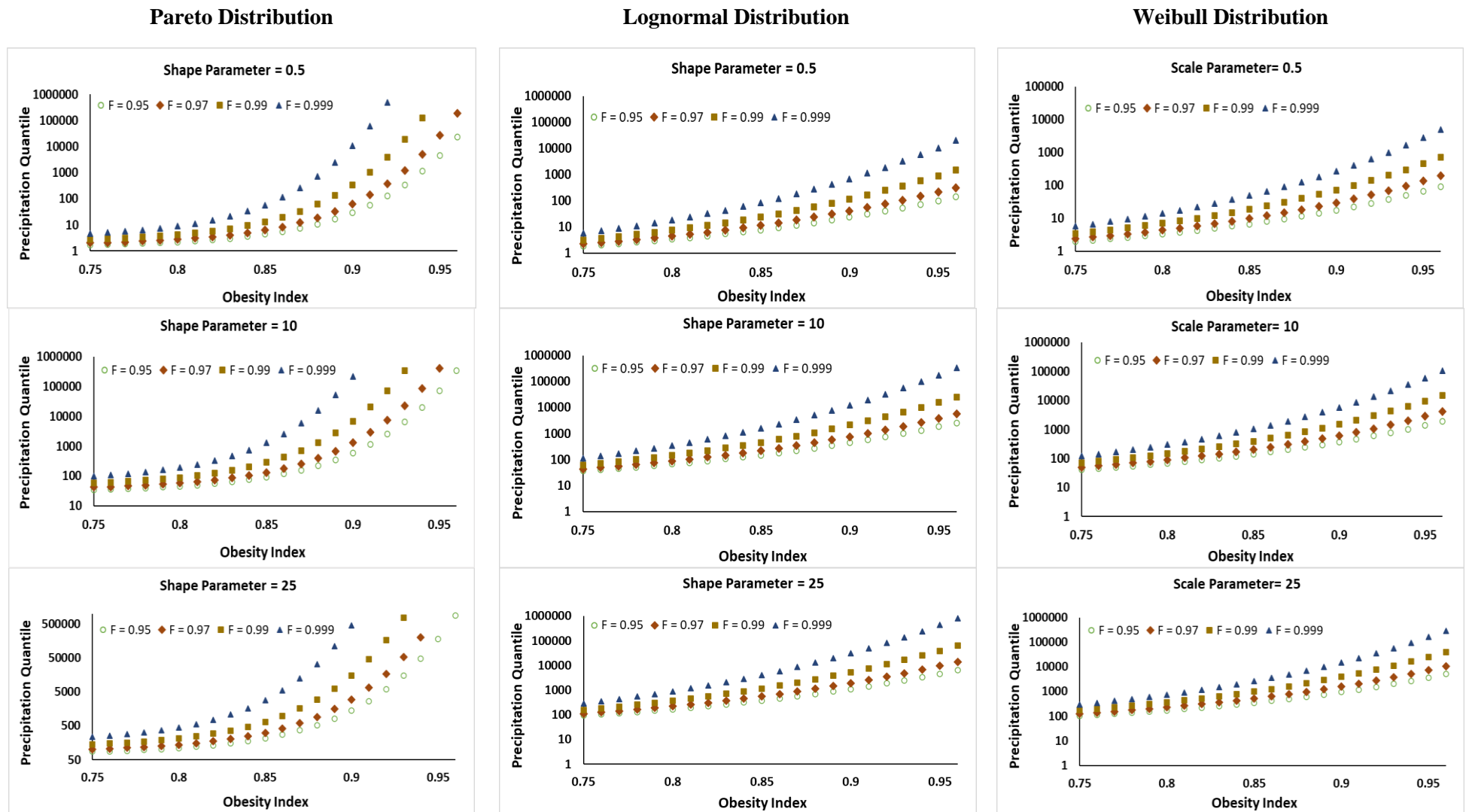
Figure 4.2 Plots showing the sensitivity of precipitation quantiles corresponding to various non-exceedance probabilities (F= 0.95, 0.97, 0.99, and 0.999) to the variation in obesity index

It can also be deduced from the figure that the non-exceedance probability corresponding to a particular value of precipitation decreases from 0.999 to 0.95 with a small increase in OB. This indicates that the frequency of occurrence of a precipitation event increases considerably with an increase in the OB, i.e., heavier tails of the probability distribution of precipitation. Therefore, it can be inferred that the OB can be useful to discriminate between the tails of heavy-tailed distributions, although the range of the OB for heavy tails is very small (i.e., from 0.75 to 1).

4.2.2   *Selection of the optimum number of random sampling*

In the case of threshold-based approaches, the analysis is generally performed on precipitation data above an arbitrary threshold value. This can occasionally lead to an inadequate length of data for performing the analysis (Papalexiou et al., 2013; Nerantzaki and Papalexiou, 2019). Contrary to this, the OB-based approach considers the entire time series of daily non-zero precipitation, which ensures the availability of a sufficient length of data for analysis. However, to apply the algorithm for the determination of the OB, information on the optimum number of random sampling (denoted by Max_count in the algorithm) is required. A simulation study was performed to address this, and the sample size N of the generated samples from the distributions was varied from 100 to 50,000. Here, we have varied Max_count from 100 to 1,00,000. Subsequently, obesity indices were determined for each of the probability distributions following the steps in the algorithm. We studied the variation of the OB with respect to the number of random sampling and the sample size for each of the distributions. Based on the analysis performed in the previous subsection, the parameters of each distribution, except the Pareto type II distribution, are chosen in such a way that both heavy- and light-tailed cases of the distributions are selected for the analysis. For the light- (heavy-) tailed case, the scale and shape parameters considered for the analysis are 2 (0.6) and 5 (0.5), respectively, based on inferences from Figure 4.1. Graphs were prepared to study the variation of the OB with respect to the number of random sampling and the size of the sample (see Figure 4.3). Details about the variation of the OB with respect to various sample sizes and the number of random sampling considered in the simulation study are provided in Tables S2 to S8 of the supplementary material of Gupta and Chavan (2022). It can be inferred from the figure and tables that a consistent estimate (i.e., fairly constant) of OB is observed for a sample size greater than 1000 and a number of random sampling greater than 5000. Hence, it can be recommended from the simulation study that sample size and number of random sampling greater than 1000 and 5000, respectively, are appropriate to estimate OB for a sample.

(i) Weibull Distribution (Shape Parameter- 0.6, Scale Parameter- 0.5)
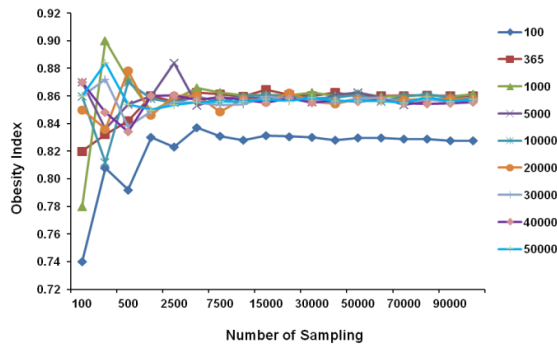
(ii) Weibull Distribution (Shape Parameter- 2, Scale Parameter- 5)
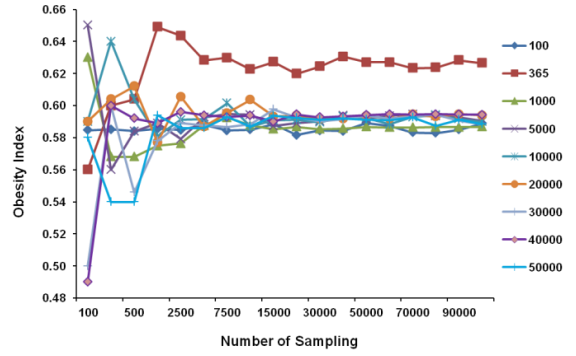
(iii) Gamma Distribution(Shape Parameter- 0.6, Scale Parameter- 0.5)
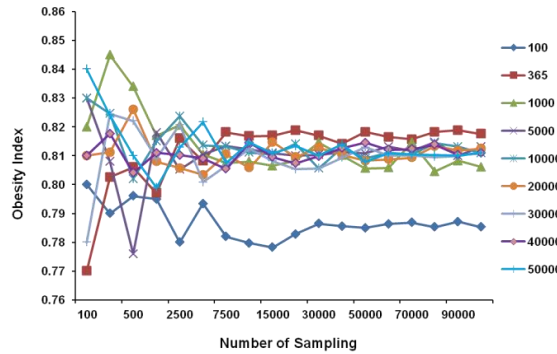
(iv) Gamma Distribution (Shape Parameter- 2, Scale Parameter- 5)
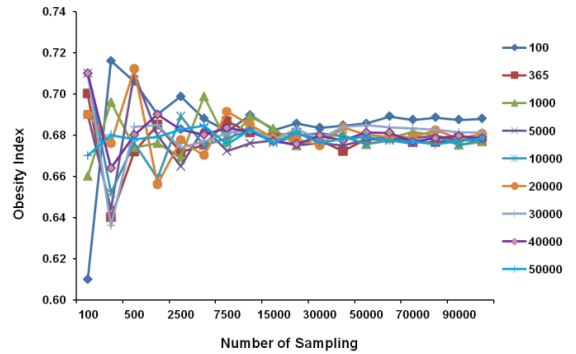
(v) Lognormal Distribution (Shape Parameter- 0.6, Scale Parameter- 0.5)
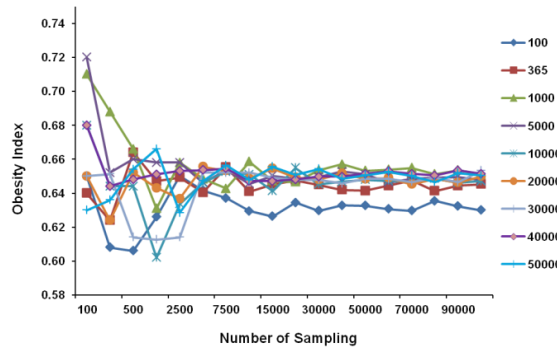
(vi) Lognormal Distribution (Shape Parameter- 2, Scale Parameter- 5)
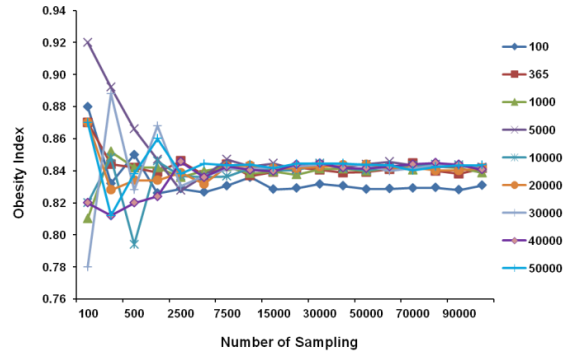
(vii) Pareto-II Distribution (Shape Parameter- 1, Scale Parameter- 5)
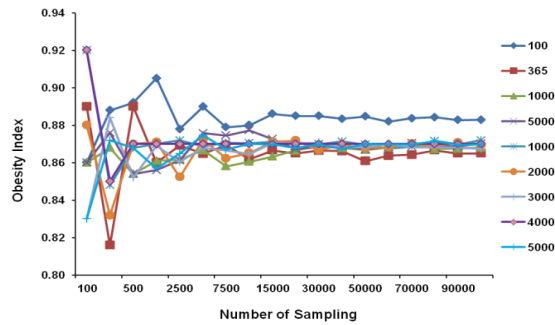
Figure 4.3 Variation of Obesity Index with respect to the number of random sampling and size of the sample for different distributions considered in this study

57

4.3   *Study area and Data used*

In this chapter, daily gridded precipitation data having a resolution of 0.25° x 0.25° prepared by the India Meteorological Department (IMD) are used for the analysis. The gridded rainfall data product was prepared by Pai et al. (2014) based on a varying network of 6955 rain gauge stations from 1901 to 2010. A record length of 53 years (1951-2004) for each grid is used for the analysis. The sample size of non-zero daily precipitation ranging from 1200 to 16000 for nearly 4949 grids over India was considered. In this study, the characterization of tails of daily precipitation was also performed at a regional scale. There are 36 Meteorological Subdivisions in India (Guhathakurta and Rajeevan, 2008). Thirty-four conterminous subdivisions were considered in the present study (see Figure 3.1 of Chapter 3).

Future projections of daily gridded precipitation data corresponding to two Representative Concentration Pathways (RCP) scenarios, namely RCP4.5 (mid-range emissions) and RCP8.5 (high-end emissions) scenarios from the year 2006 to the year 2099, were obtained from the Coordinated Regional Climate Downscaling Experiment-South Asia (CORDEX-SA), using 6 Regional climate model (RCM) experiments which are a part of the initiatives of the World Climate Research Programme (WCRP) (http://www.cordex.org/) (Taylor et al., 2012; Sanjay et al., 2017; Nikulin et al., 2011; Choudhary et al., 2017). Details on the CORDEX-SA experiment can be found in section S2 in the supplementary material of Gupta and Chavan, 2022. Further, the details related to the 6 RCMs considered in this study are presented in Table 4.1.

RCMs are derived by dynamically downscaling the large-scale climate variables of GCM simulations to yield spatially and physically consistent outputs (Sharma et al., 2011; De Sales and Xue, 2011). However, the RCMs can have considerable biases with respect to observed data due to different spatial resolution, model systematic error, and inaccurate physical parameterizations (Maurer and Hidalgo, 2008; Christensen et al., 2008; Ghosh and Mujumdar, 2009; Chen et al., 2013; Turco et al., 2013). Thus, the daily precipitation data from both IMD gridded data and CORDEX-SA RCMs for historical periods (1951 to 2004) was checked for the presence of any bias.

Table 4.1 Details of CORDEX-SA experiments considered in the present study (source: CORDEX South-Asia Database, CCCR, IITM; http://cccr.tropmet.res.in/cordex/files/downloads.jsp)

No	Experiment Name	Name Used	RCM Description	Driving GCM	Contributing CMIP5 Modeling Center
1	ACCESS-CSIRO CCAM	ACCESS		ACCESS1.0	CSIRO, Australia
2	CNRM-CM5-CSIRO CCAM	CNRM		CNRM-CM5	Centre National de RecherchesMe´te´orologiques (CNRM), France
3	CCSM4-CSIRO CCAM	CCSM4	Commonwealth Scientific and Industrial Research Organisation (CSIRO), Conformal Cubic Atmospheric Model (CCAM; McGregor and Dix, 2001)	CCSM4	National Center for Atmospheric Research (NCAR), USA
4	GFDL-CM3-CSIRO CCAM	GFDL-CM3		GFDL-CM3	National Oceanic and Atmospheric Administration (NOAA), Geophysical Fluid Dynamics Laboratory (GFDL), USA
5	MPI-ESM-LR CSIRO-CCAM	MPI		MPI-ESM-LR	MPI-M, Germany
6	NorESM1-M-CSIRO CCAM	NorESM1		NorESM1-M	Norwegian Climate Centre (NCC), Norway

The box plots presented in Figure 4.4 show the comparison between RCM-based daily precipitation and observed precipitation data for (i) a typical IMD grid (latitude 14.5° N and longitude 74.5° E) and (ii) Arunachal Pradesh Subdivision. Herein, the bias is defined as the difference in the quartiles (i.e., 5th, 25th, 50th, 75th, and 95th percentile) and interquartile range (i.e., the difference between 75th and 25th percentiles). The issue of bias in daily precipitation estimates was resolved through the additive multiple change factor method following Semadeni-Davies et al. (2008) and Anandhi et al. (2011). The description of the method is provided in section S3 of the supplementary material of Gupta and Chavan (2022).

(i)

Box Plot for Daily rainfall

(ii)

Box Plot for Daily rainfall (Arunachal Pradesh)
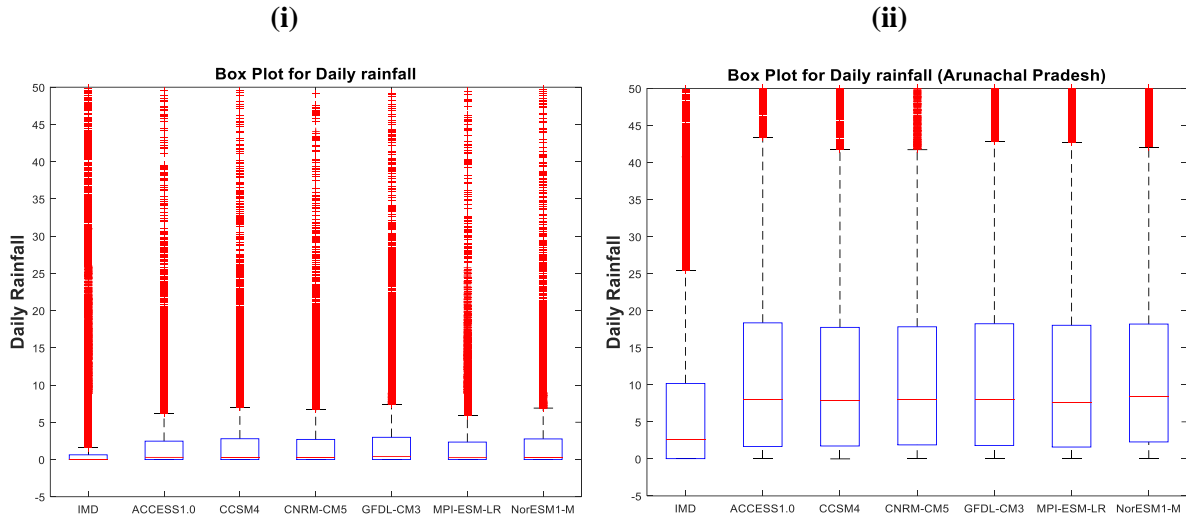
Figure 4.4 Box plot showing the comparison between RCM based daily precipitation and observed precipitation data for (i) a typical IMD grid (latitude 14.5° N and longitude 74.5° E) and (ii) Arunachal Pradesh Subdivision

4.4 Results

In this section, we characterized the tail behavior of daily gridded precipitation over India by using the OB-based approach and threshold-based approaches. Details regarding threshold-based approaches can be found in Appendix A.

4.4.1 Characterization of tail behavior of daily gridded precipitation using obesity index

The OB-based approach was utilized to characterize the tails of daily gridded precipitation data over India. Following the algorithm proposed in section 4.2, we evaluated the OB at each grid over India. The sample size (N) available for the analysis was always greater than 1000. The number of random sampling considered for the analysis was 5000. Figure 4.5 shows the variation of the OB over India. The estimates of OB for grids range from 0.76-0.93. It can be observed that all the grids over India possess OB values greater than 0.75. Hence it can be concluded that the non-zero daily gridded precipitation data over India exhibit heavy tails. The dominance of heavy-tailed behavior of daily precipitation data indicates that the extreme precipitation events over India have a high probability of occurrence.

The OB values in this study varied from 0.85 to 0.95 for the western parts of Gujarat, Rajasthan, Punjab, North-eastern parts of Ladakh, and the coastal part of Tamil Nadu and Andhra Pradesh, signifying frequent occurrence of extreme precipitation events. The index varied from 0.75 to 0.85 for the North Eastern States of India, Northwestern parts of Jammu & Kashmir, Himachal, Uttrakhand, parts of the west-central region, and the west coast region of Maharashtra.

60
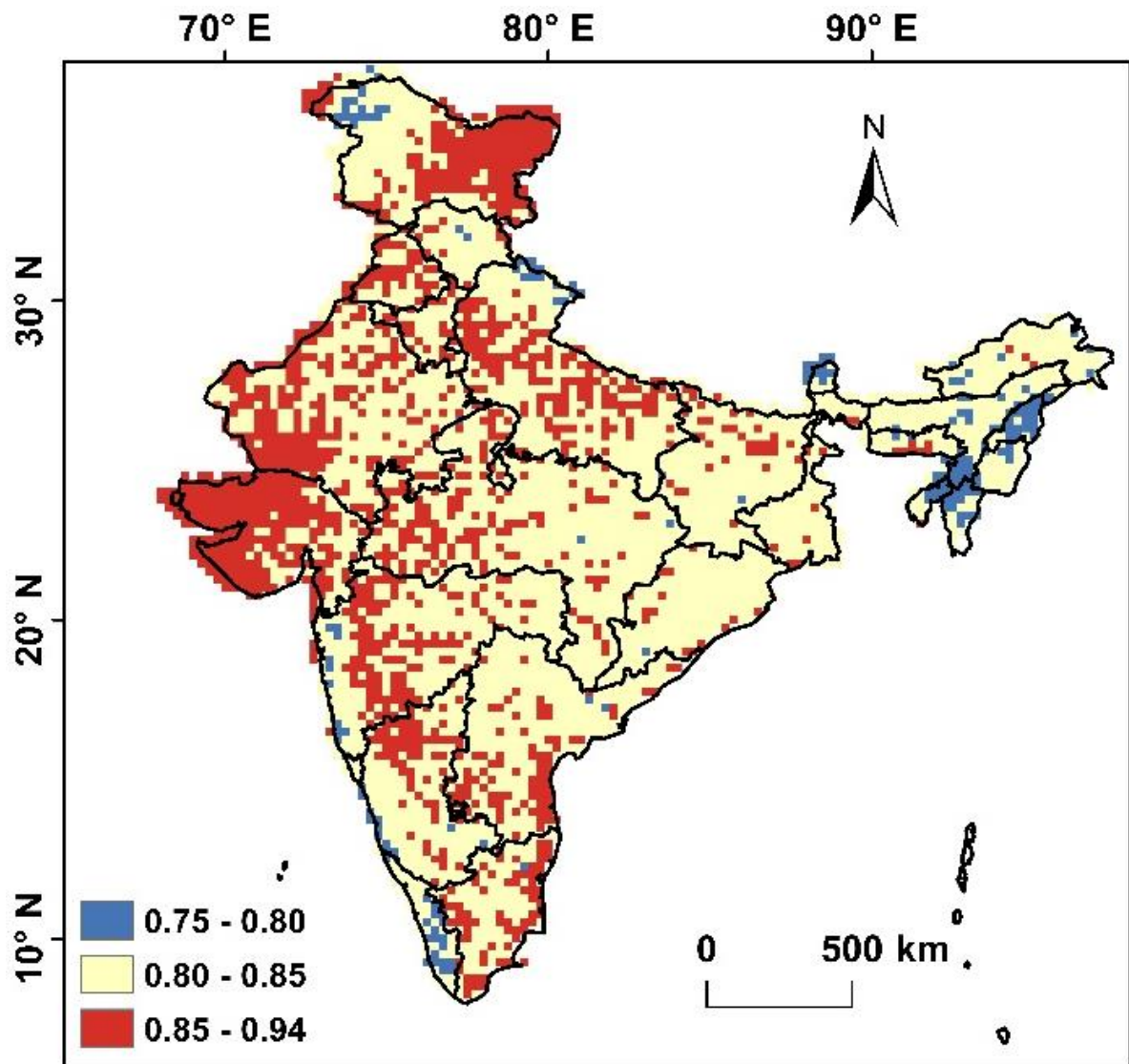
Figure 4.5 Spatial variation of Obesity index over India for IMD daily gridded precipitation data for the historical period (1951-2004)

Similar findings over these regions were observed by previous studies (e.g., Goswami et al., 2006; Dash et al., 2009; Ajayamohan and Rao, 2008; Guhatakurta et al., 2011; Vittal et al., 2013; Mishra et al., 2014; Krishnan et al., 2016; Roxy et al., 2017; Sarkar and Maity, 2020). The high frequency of extreme precipitation in a different part of the country has been associated with various global and local causes like the abrupt global change of the climatic system caused by a regime shift in the 1970s in various climatic factors like the Arctic Oscillation (AO), East Asian summer monsoon (EASM), East Asian winter monsoon (EAWM), El Niño–Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), Aleutian low (AL), Pacific decadal oscillation (PDO), Western Pacific subtropical high (WPSH) and Indian summer monsoon rainfall (ISMR) (Biondi et al., 2001; Chowdary et al., 2006; Zhou et al., 2009; Vittal et al., 2013; O'Kane et al., 2014; Chen et al., 2015; Sabeerali et al., 2015; Sahana et al., 2015; Vinnarasi and Dhanya, 2016; Dai et al., 2018) or some local changes such as Urbanization (Singh et al., 2014; Shastri et al., 2015). Along the west coast, the southern part of the coast (e.g., Kerela) has

relatively lighter tails (i.e., OB ranging from 0.75 - 0.80) as compared to the northern parts of the west coast (Konkan and Goa) for the observed period. This can be attributed to the change in regional precipitation behavior (such as the sporadic nature of rainfall) and climatic patterns as observed by Varikoden et al. (2019).

4.4.2 Characterization of tail behavior of daily gridded precipitation using threshold-based approaches

The threshold-based approaches proposed by Nerantzaki and Papalexiou (2019) and Papalexiou et al. (2013) were used to characterize the tails of daily precipitation data over India. Based on Nerantzaki and Papalexiou (2019) approach, the estimate of the MEF slope was tested against the null hypothesis of zero slope indicating exponential tail by considering 90% or 95% confidence interval. For brevity, typical plots between Threshold Precipitation *vs.* Empirical MEF helpful in identifying the sub-exponential tail (i.e., large positive slope) and exponential tail (i.e., close to zero slope) of the probability distribution of daily precipitation are illustrated in Figure 4.6.

(i)

(ii)
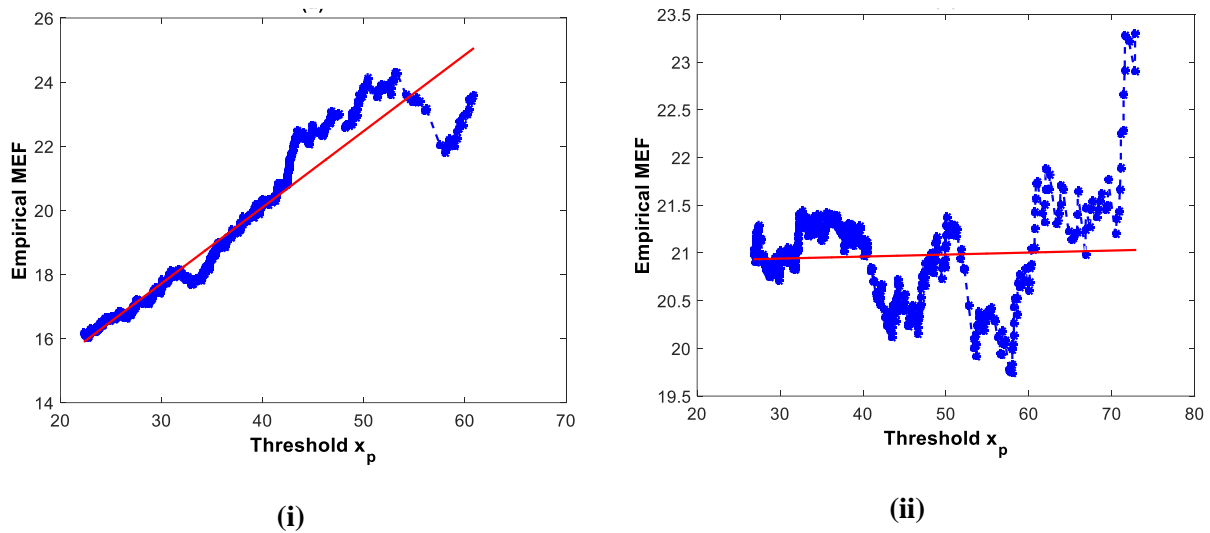
Figure 4.6 Typical plots between Threshold Precipitation *vs.* Empirical Mean Excess Function for identifying (i) sub-exponential tail and (ii) exponential tail of probability distribution of daily precipitation. For case (i), the value of the slope is found to be 0.24 with its 90% Confidence Interval as (−0.07, 0.07), whereas, for case (ii), the value of the slope is found to be 0.002 with its 90% Confidence Interval as (−0.09, 0.08). Coordinates of the grids analyzed in (i) and (ii) are (latitude 8.5° N and longitude 77° E) and (latitude 19° N and longitude 79° E), respectively.
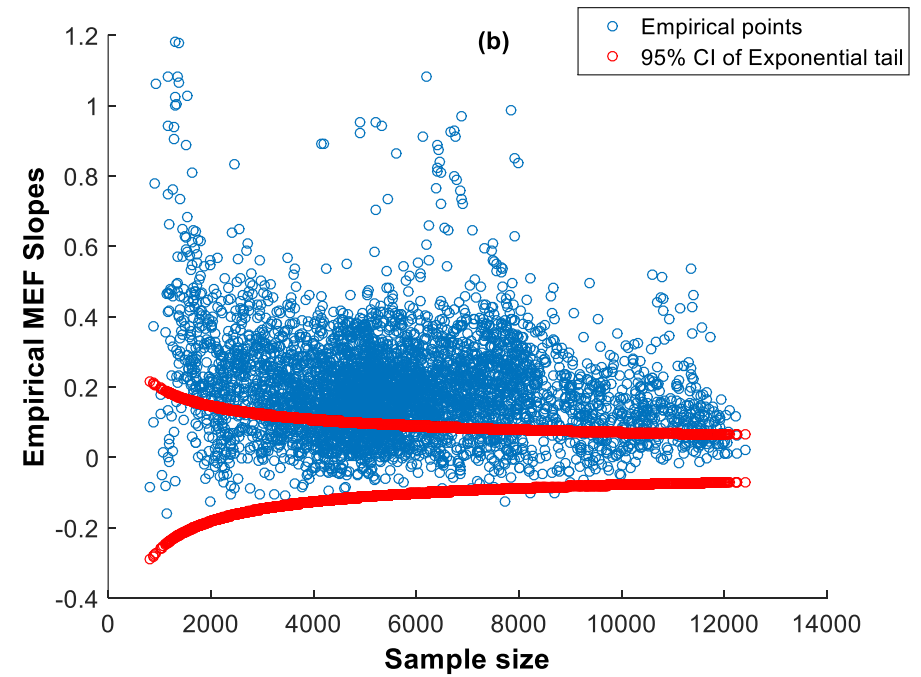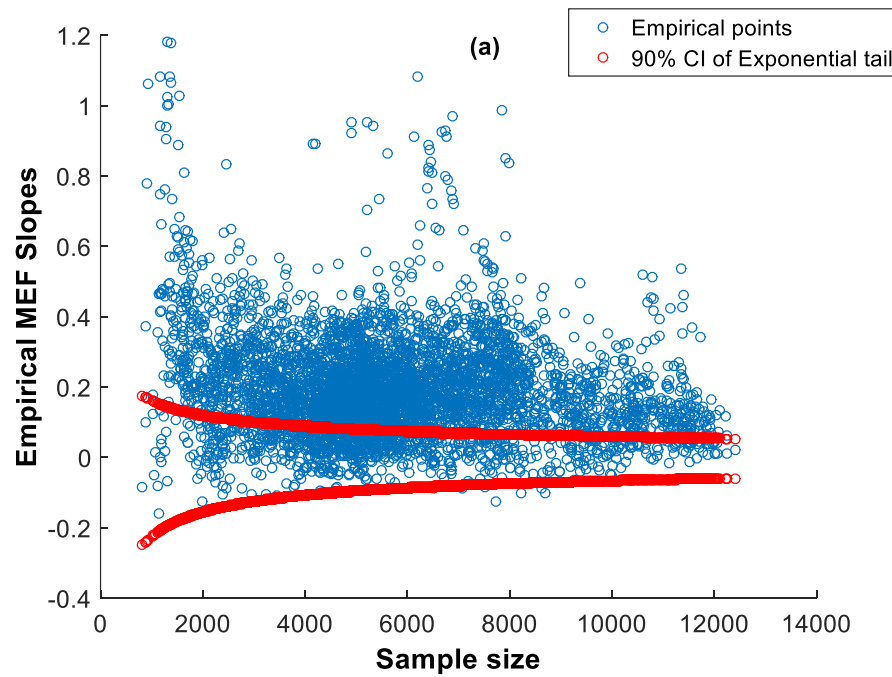
Figure 4.7 Testing for significant slopes of plots between Threshold Precipitation *vs*. MEF for 4949 grids over India considering (i) 90% confidence interval and (ii) 95% confidence interval against zero slope (i.e., case of exponential tail)

We found that the hypothesis of the presence of an exponential tail is rejected for nearly 80.62% and 75.95% of grids over India at 90% and 95% confidence intervals, respectively (refer to Figure 4.7). Further, analysis based on a one-tailed test revealed that nearly 99.89% grids out of 80.62% grids and 99.79% grids out of 75.95% grids possess sub-exponential tails at 5% and 10% significance levels, respectively. Figure 4.8 shows the spatial distribution of the absolute values of the 4949 observed MEF slopes of daily precipitation over India.
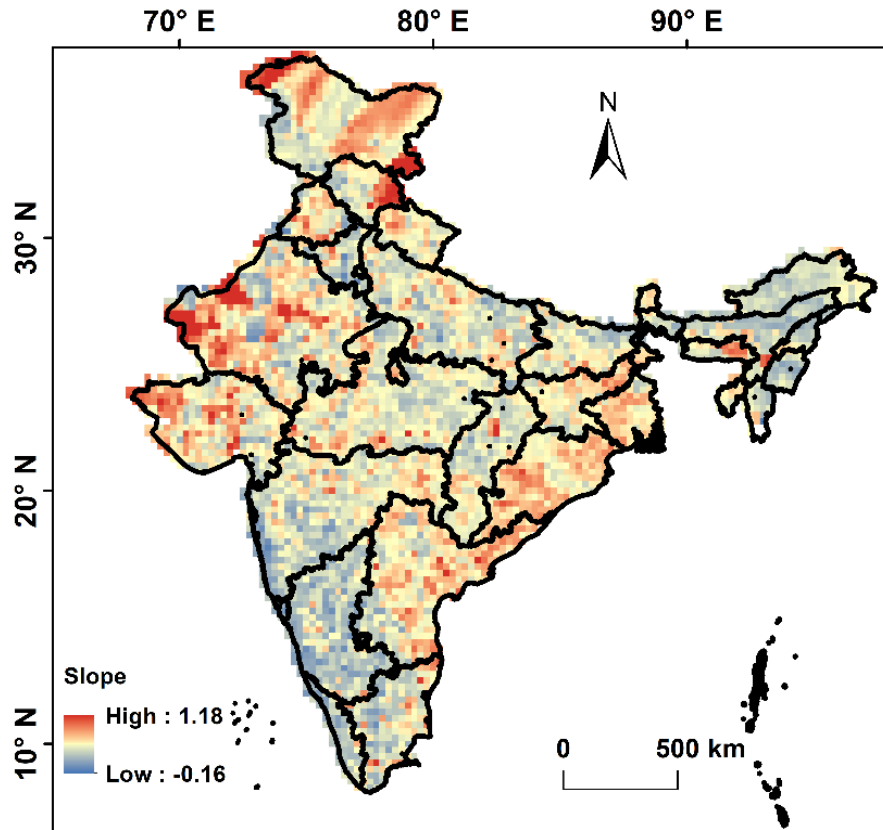
Figure 4.8 Spatial distribution of the absolute values of the 4949 observe MEF slopes of daily precipitation over India

Figures 4.9 (i, ii, iii) represents the spatial distribution of sub-exponential and exponential/hyper-exponential tails over the country assessed through threshold-based approaches given by Nerantzaki and Papalexiou (2019) and Papalexiou et al. (2013). Since very few grids showed the hyper-exponential tails, they were merged with the exponential class for representation. The figures indicated that the majority of grids over India show the presence of heavy tails for daily precipitation with some lighter tails over the West Coastal region, Southern parts Karnataka, Northern part of Tamil Nadu, and some parts of Kerala, Maharashtra, and the North Eastern States of India.

(i)

(ii)

(iii)
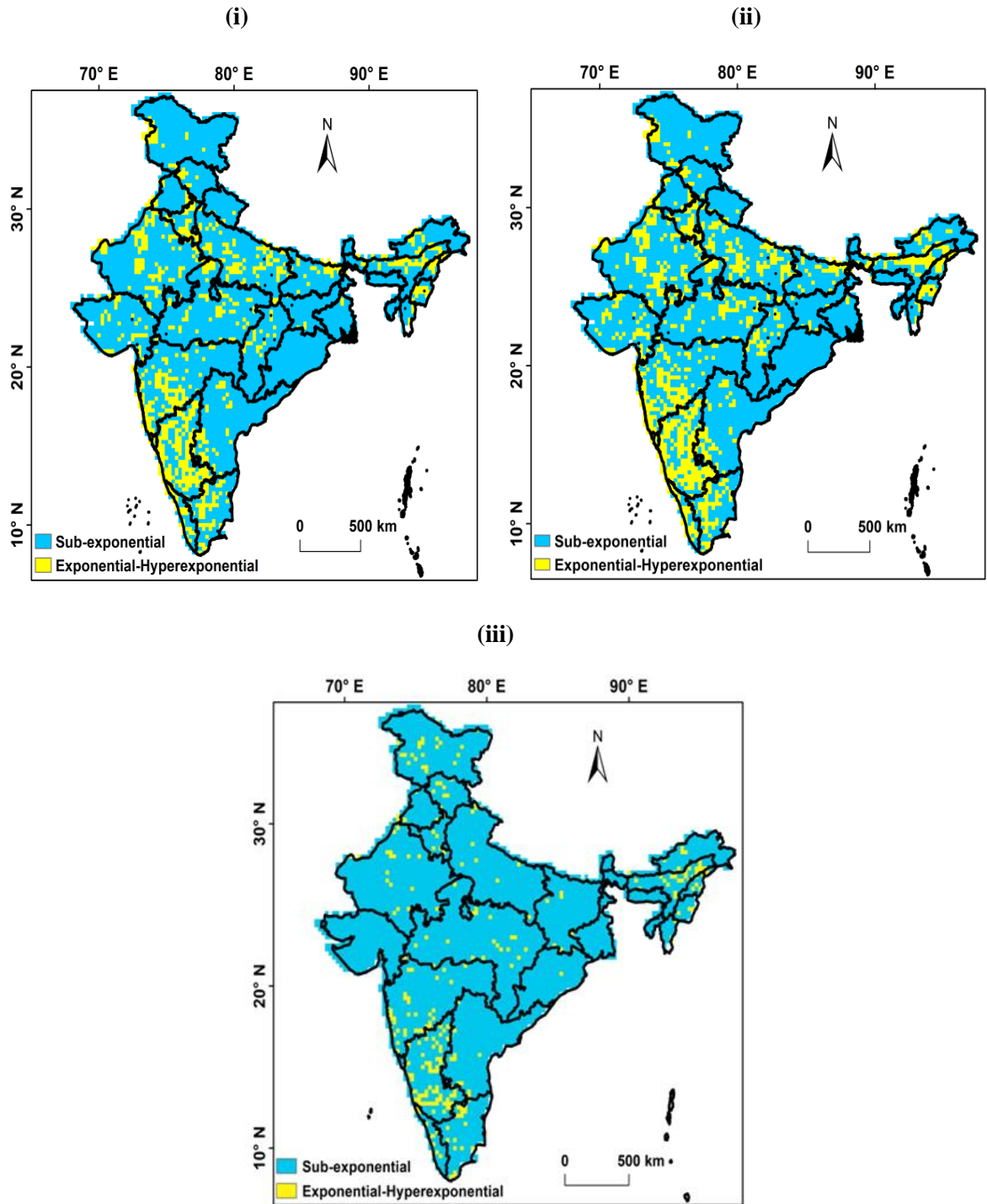
Figure 4.9 Spatial distribution of sub-exponential and exponential–hyperexponential tails over India obtained based on Nerantzaki and Papalexiou (2019) at (i) 90% confidence interval, (ii) 95% confidence interval, and (iii) shows the spatial distribution of tails obtained based on Papalexiou et al. (2013)

In the case of the approach given by Papalexiou et al. (2013), the decision on the heaviness of the tails of daily precipitation data was based on fitting four probability distributions to the precipitation data in the tail part. Parameters of fitted distributions are presented in Table S10 in the supplementary material of Gupta and Chavan (2022). Figure S3 in the supplementary material of Gupta and Chavan (2022), shows the best-fitted distributions for representing tails of daily precipitation data over India.

The figure does not reveal any regular pattern for best-suited distributions. Thus, the distributions were divided into sub-exponential and exponential-hyper-exponential groups, as shown in Figure 4.9 (iii). The former group comprised Pareto Type II distribution, Lognormal distribution, and Weibull distribution with $\alpha < 1$, whereas the latter group included Gamma distribution and Weibull distribution with $\alpha > 1$ distribution. It was found that most of the daily precipitation data over India possess heavy-tailed distribution. Table 4.2 represents the summary of the best-suited probability distributions for daily precipitation over India.

Table 4.2 Summary of the best-fitted probability distributions for daily precipitation over India

Distribution considered for Tail Fitting	Number of grids (in %)
Lognormal	43.69
Pareto	33.83
Weibull	16.75
Gamma	5.74

It was observed that the gamma distribution, which is considered the most appropriate distribution for representing daily precipitation, was the least-suited probability distribution. The lognormal distribution is found to be the most suitable probability distribution for representing tails of daily precipitation data at 43.69% grids over India, followed by Pareto distribution at 33.83% grids. Further, the probability distributions ranked based on the minimum PRMSE criterion as shown in Figure S4 of the supplementary material of Gupta and Chavan, 2022. Lognormal secured the lowest rank, followed by Pareto and Weibull distribution in that order. The gamma distribution was ranked the highest and was considered as the worst performer for representing daily precipitation tails over India. Additionally, the suitability of probability distributions was also tested by considering their fits in pairs or couples, i.e., Lognormal *vs.* Pareto; Pareto *vs.* Weibull; Pareto *vs.* Gamma; Lognormal *vs.* Weibull; Lognormal *vs.* Gamma and Weibull *vs.* Gamma. Figure S5 in the supplementary material of Gupta and Chavan (2022), shows the results for fits in couples. It was once again observed that the Lognormal distribution was better suited than any of the three probability distributions. Pareto came out to be the second best fit, better than Weibull and light-tailed Gamma distributions. It should be noted that the Weibull distribution predominantly tends to have $\alpha < 1$, which is a heavy-tailed distribution. Amongst the 4949 grids, subexponential tails were best fitted in 93.89% of the grids, while the remaining 6.11% grids were found to have exponential-hyperexponential tails.

4.4.3   *Comparative analysis between the OB-based approach and threshold-based approaches*

All three approaches seem to provide consistent inference regarding heavy-tailed behavior in many parts of the country. The relative assessment of the results revealed that around 3871 grids (i.e., 78.21%) showed heavy-tailed behavior based on all three approaches. The assessment performed in

pairs revealed that approaches based on Papalexiou et al.(2013) and OB indicated the presence of heavy-tailed behavior at 4647 grids (i.e., 93.89%), whereas similar behavior of tails was observed at 3853 grids (i.e., 77.85%) when the approaches based on Nerantzaki and Papalexiou (2019) and OB were considered. In the case of threshold-based approaches, similar tail behavior was found at 3871 grids (i.e., 78.21%). Wherever light-tailed behavior is inferred by the threshold-based approaches (e.g., coastal parts of Maharashtra, Karnataka, and Kerala; parts of North East States), the OB-based approach provides comparatively lighter tails (i.e., OB ranging from 0.75 to 0.8) in those regions (as compared to rest of India) (please refer to Figure 4.5 and Figure 4.9). However, the three approaches might not always provide similar assessments as they inherently differ from each other in the way they analyze the tail behavior of daily precipitation data. The regions belong to various climate zones, and the local physical process can influence the tail in those regions (e.g., parts of peninsular India and some parts of north India). The differences observed in the case of threshold-based approaches, especially in the Northcentral part of India, can be attributed to strategies used for demarcating the tail of daily precipitation data. In the case of Papalexiou et al. (2013), an annual exceedance series (AES) was used for demarcating the tail of daily precipitation data, while the demarcation of the tail was accomplished by fixing the 90th (lower limit) and 99th (upper limit) percentile of the data in the case of Nerantzaki and Papalexiou (2019). This shows that the difference in the strategy to fix some threshold can lead to dissimilar conclusions (Kiran and Srinivas, 2021). Contrary to the threshold-based approaches, the OB-based approach considers the entire probability distribution of daily precipitation data for inferring the tail behavior.

4.4.4 *Characterization of tail behavior of future projections of daily gridded precipitation*

In this study, future projections of daily gridded precipitation from the year 2006 to the year 2099) from 6 RCMs of the CORDEX-SA experiments corresponding to two scenarios (RCP4.5 and RCP8.5) were considered. Figure 4.10 shows the spatial distribution of OB estimates determined for the future climate scenarios (i.e., RCP4.5 and RCP8.5). Visual interpretation of the figure illustrates that the future daily precipitation data exhibits heavy-tailed distributions over the entire India except for the southern region of Tamil Nadu State, where light-tailed distributions were observed in the case of some RCMs (GFDL-CM3, NorESM1-M, ACCESS 1.0). It was observed that the tails of future projections of daily precipitation pertaining to the RCP4.5 scenario for all the RCMs tend to be lighter than those corresponding to RCP8.5. A summary of the percentages of grids having $OB < 0.75$, $0.75 < OB <$ to 0.85, and $OB > 0.85$ are provided in Table 4.3. It can be noted that the percentage of grids with $OB > 0.85$ $(0.75 < OB < 0.85)$ is higher (lower) for the RCP8.5 scenario as compared to the RCP4.5 scenario except for CNRM-CM5. Overall, this analysis reveals that the frequency of occurrence of extreme precipitation events based on the RCP8.5 scenario is higher than the RCP4.5 scenario. This finding is in line with many studies in the past studies (e.g., Mukherjee et al., 2018; Gusain et al., 2019; Rai et al., 2019; Rao et al., 2020).

Name of RCMs	RCP4.5	RCP8.5
ACCESS 1.0		
CCSM4		
CNRM-CM5		
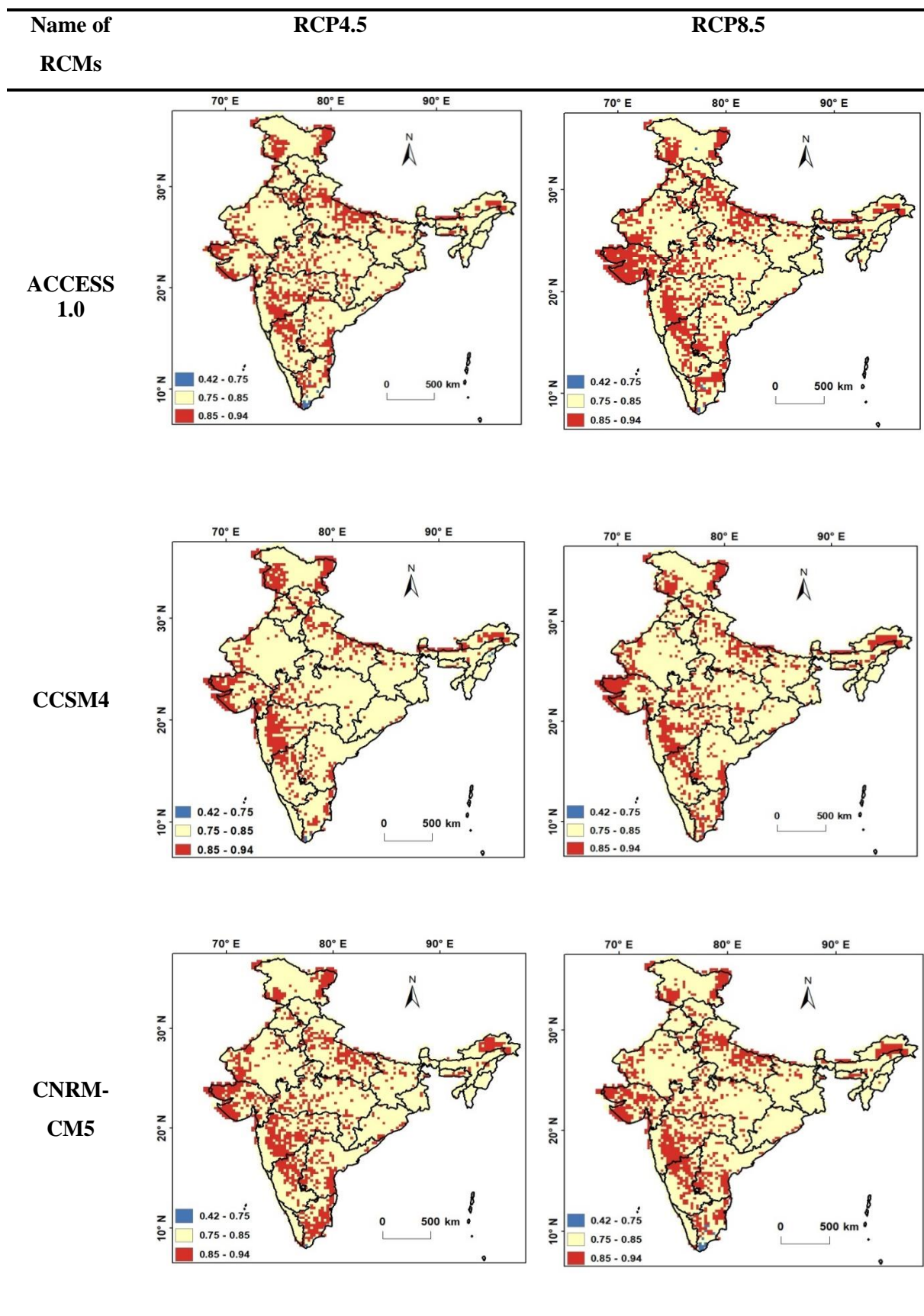
Figure 4.10 Spatial distribution of obesity index under RCP4.5 and RCP8.5 scenarios obtained from 6 RCMs for the future time period (2006-2099)

Name of RCMs	RCP4.5	RCP8.5
GFDL-CM3		
MPI-ESM-LR		
NorESM1-M		
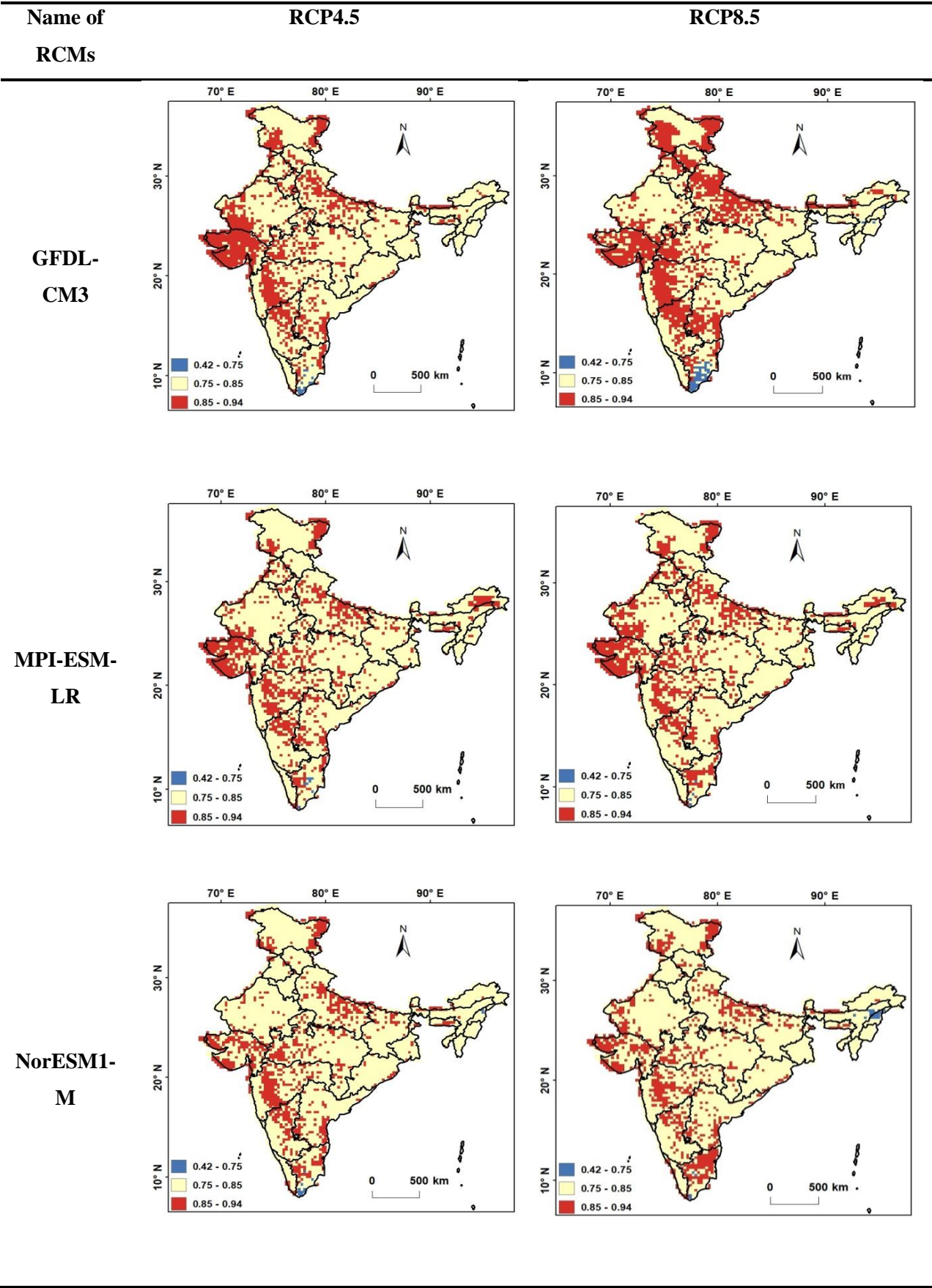
Table 4.3 Percentage of grids having obesity indices less than 0.75, 0.75 to 0.85, and greater than 0.85 for future projections of daily precipitation over India obtained from 6 RCMs

RCM	Percentage of Grids		
	OB<0.75	0.75<*OB*<0.85	*OB*>0.85
ACCESS1.0	0.22	76.49	23.29
	0.12	71.52	28.36
CNRM-CM5	0.00	74.14	25.86
	0.24	75.22	24.54
CCSM4	0.12	79.61	20.27
	0.04	76.88	23.08
GFDL-CM3	0.30	71.74	27.96
	1.03	65.50	33.47
MPI-ESM-LR	0.16	75.25	24.58
	0.08	73.41	26.51
NorESM1-M	0.26	78.62	21.12
	0.49	78.00	21.51

4.4.5   *Characterization of daily precipitation over meteorological subdivisions of India*

We extended the utility of the concept of OB to investigate the heaviness of tails of daily precipitation data at a regional scale over India. The estimate of OB for each subdivision was determined by taking the average of the indices estimated for individual grids falling within that subdivision boundary. The OB estimates for the subdivisions were determined for both historical and future time periods. Figure 4.11 illustrates the obesity indices for 34 conterminous subdivisions over India for the historical time period. The estimates of the OB ranged from 0.8 to 0.87, which shows that the precipitation within the Subdivisions exhibits heavy-tailed behavior. Less heavy tails (indicating less frequency of occurrence of extreme daily precipitation as compared to the rest of India) were observed in North-East, Western coastal parts of India, West-central regions like Bihar, Jharkhand, Orrisa and West Bengal, and Uttranchal in North India. Nandargi et al. (2016) observed a decreasing trend in magnitude and number of extreme rainfall in 42% of Uttarakhand stations, especially after 1970.
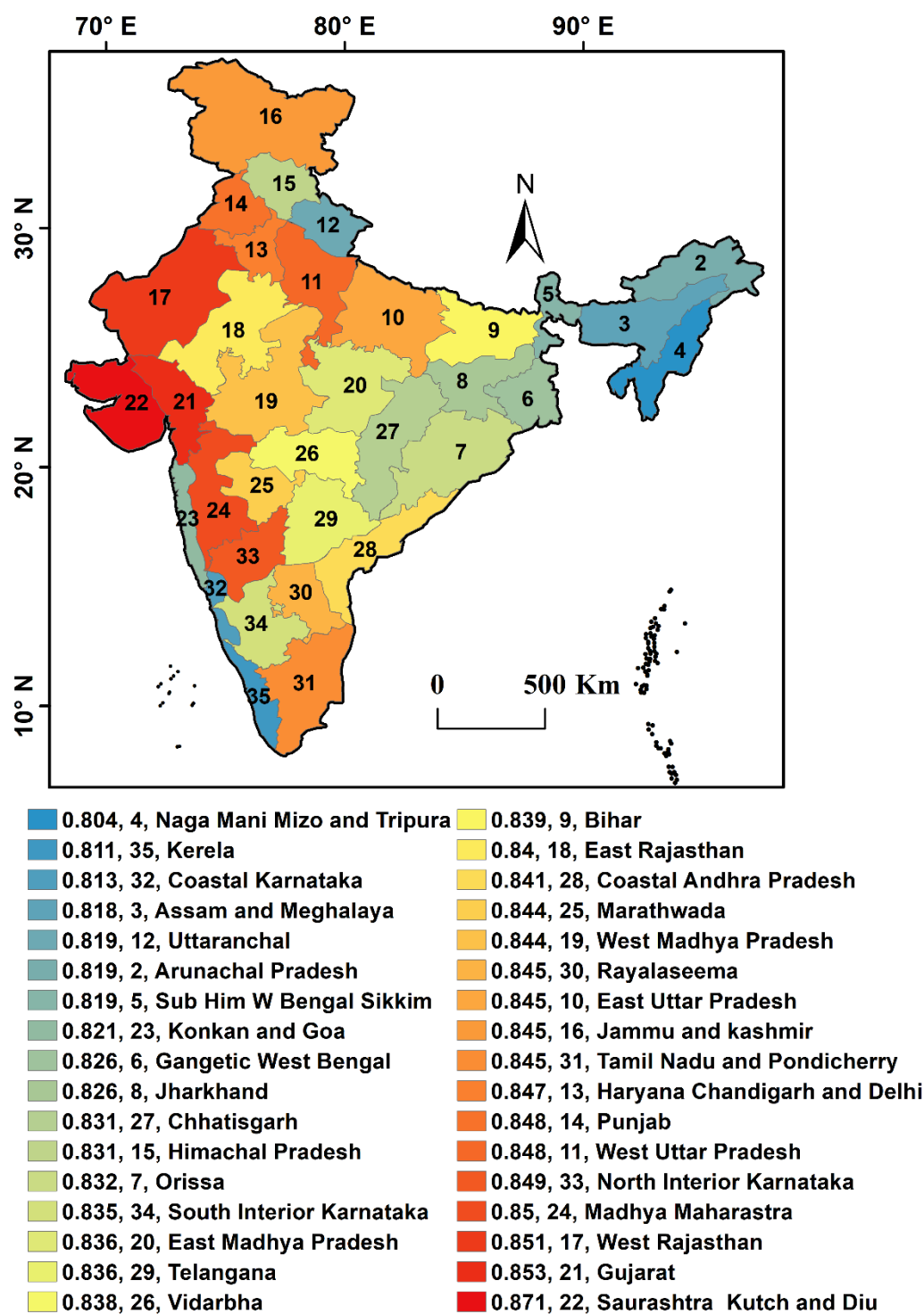
Figure 4.11 Estimates of obesity indices for 34 Meteorological Subdivisions over India for the historical time period. The first number represents the Obesity Index, second number represents the subdivision's Id. The estimates of OB given here are the mean of OBs of the gridded time series found inside each subdivision

Heaviest tails (indicating a very high frequency of occurrence of extreme daily precipitation) were found in Northwestern parts (especially Saurashtra, Kutch and Diu, Gujarat, Rajasthan, Punjab, Haryana Chandigarh, and Delhi), followed by subdivisions in west-central India (like Madhya Maharashtra,

Marathwada, Rayalseema, North Interior Karnataka) and northern central states (West Uttar Pradesh, East Uttar Pradesh). Similar observations are also made in some previous studies. Dave and James (2017) showed a significant increase in intense/extreme rainfall events (rainfall corresponding to the 95th, 98th, 99th, and 99.5th percentiles) along with the increased frequency of heavy and very heavy rainfall events over Gujarat State (India), especially in Saurashtra region. Dash et al (2009) observed significant increasing trends of heavy-intensity rainfall and found an increase in days of heavy rainfall over northwest India, northeast India, and central India. These studies are in line with our finding of the OB values for different meteorological subdivisions for the historical period.

For RCP4.5 and 8.5 scenarios, estimates of OB were determined for each of the Subdivisions. Figure 4.12 shows the estimates for 34 Meteorological Subdivisions over India for future time periods corresponding to the 6 RCMs under study. A comparative analysis was performed to investigate the changes in the OB over time, i.e., historical and future time periods (see Figure 4.13). Firstly, we examined only the change (irrespective of increase or decrease) in the mean OB estimate for each subdivision over the time periods. A Z-test at a 5% significance level was considered in this study to test the null hypothesis that the mean OB for a meteorological subdivision has not changed from historical to future periods. The changes in OB over time were not consistent across the subdivisions. Detailed observations regarding the number of RCMs showing the change in the mean OB (based on the Z-test) from historical to future time periods corresponding to RCP4.5 and 8.5 scenarios are presented in Table S12 of the supplementary material of Gupta and Chavan (2022). Further, Kolmogorov-Smirnov (KS) (Massey, 1951; Marsaglia, 2003) test was used to compare the distribution of OB corresponding to the historical time period relative to its distribution corresponding to the future RCP scenarios for each of the subdivisions at a 5% significance level. Figure S6 in the supplementary material of Gupta and Chavan (2022), presents the cumulative distribution functions (CDFs) of OB estimates corresponding to both time periods for 34 subdivisions. Again, the differences in the CDFs of OB over time were not consistent across the subdivisions (see Table S12 in the supplementary material of Gupta and Chavan (2022)). It was difficult to draw a conclusion regarding the change in OB over time from the results obtained based on different RCMs. To overcome this difficulty, an ensemble of 6 RCMs was considered for studying the climatic changes in mean OB estimate as it ensures a robust estimation of climate state by averting errors and internal variability of an individual model (Rai et al., 2019).

Name of the RCM	RCP4.5	RCP8.5
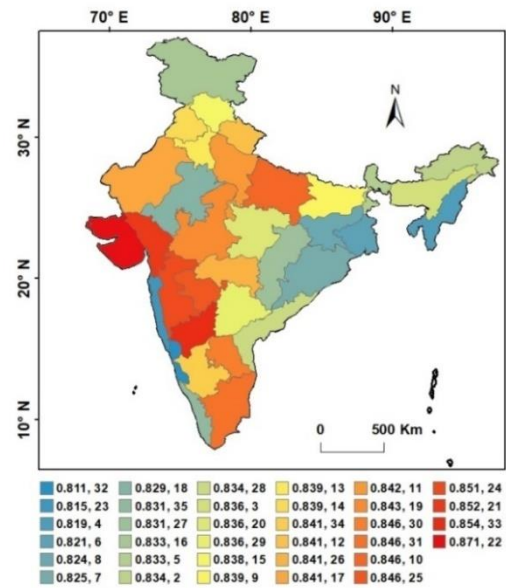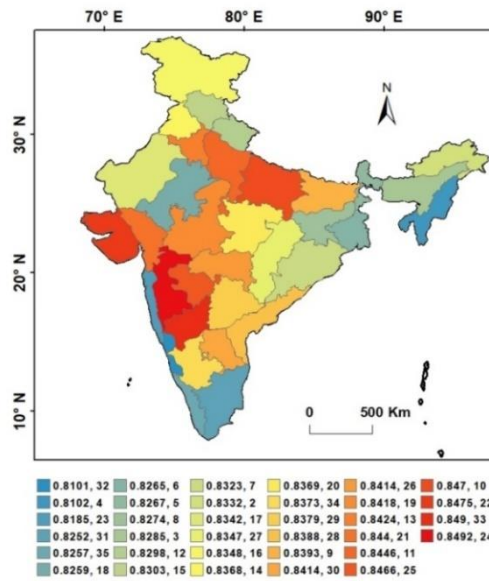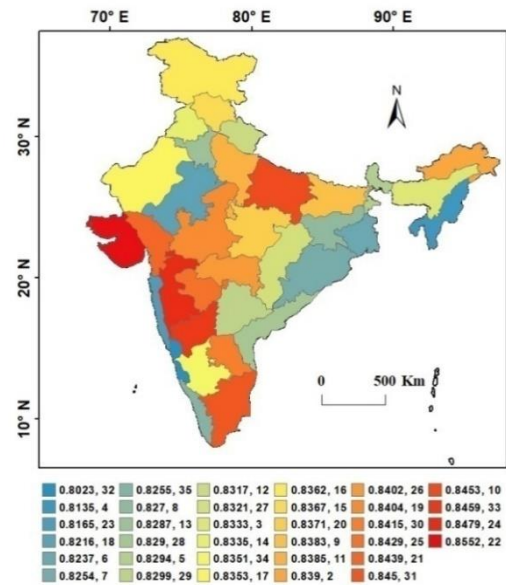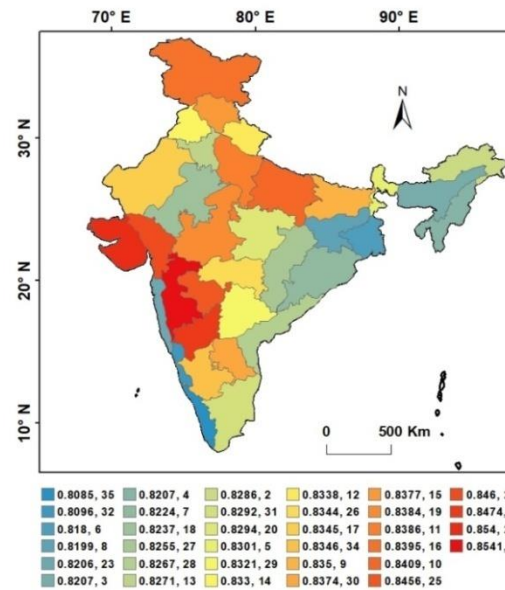
Figure 4.12 Estimates of obesity indices for 34 conterminous Meteorological Subdivisions over India for future time period (2006-2099)

Name of the RCM	RCP4.5	RCP8.5
CNRM-CM5		
GFDL-CM3		

Name of the RCM	RCP4.5	RCP8.5
MPI-ESM-LR		
NorESM 1-M		

(i)

(ii)
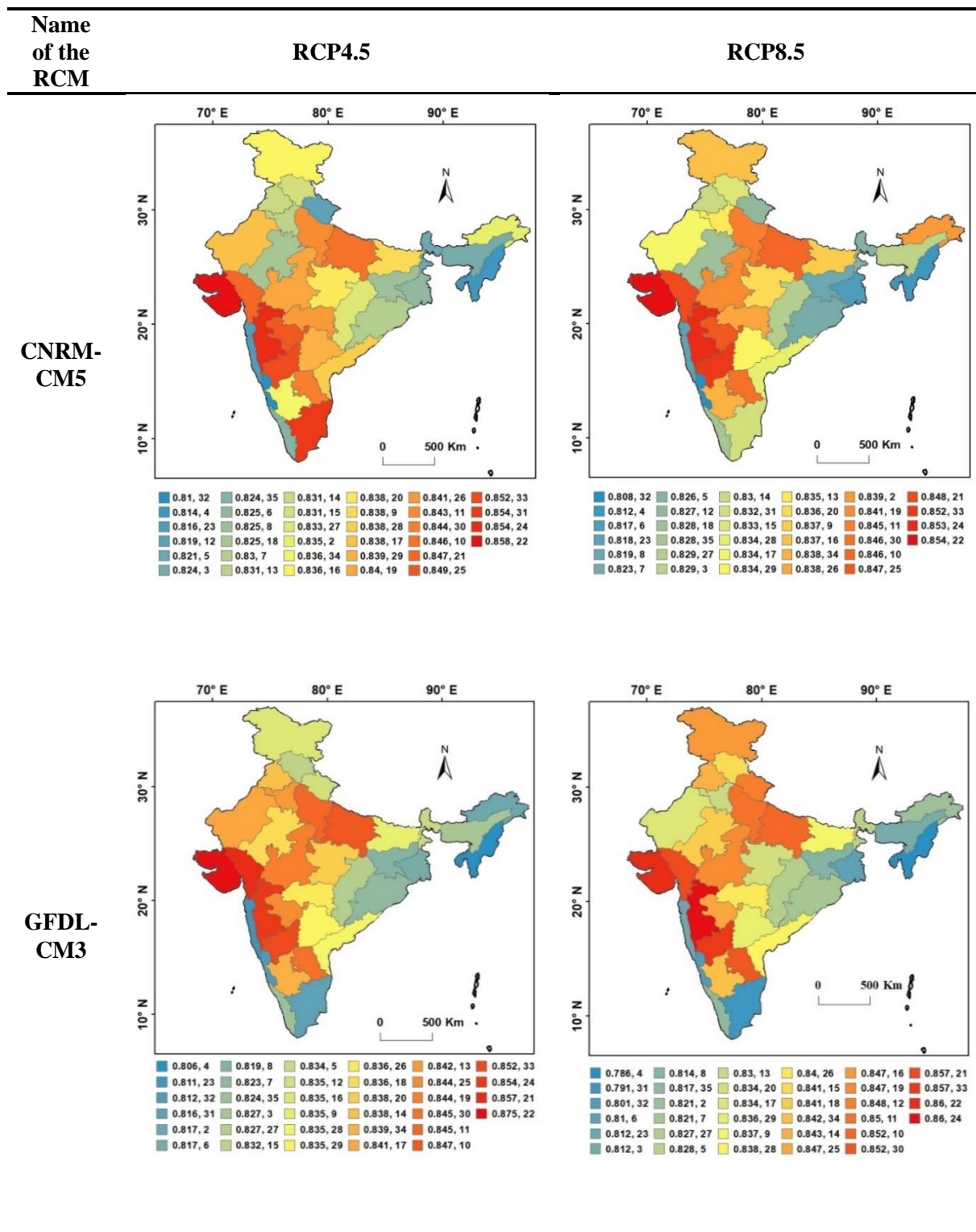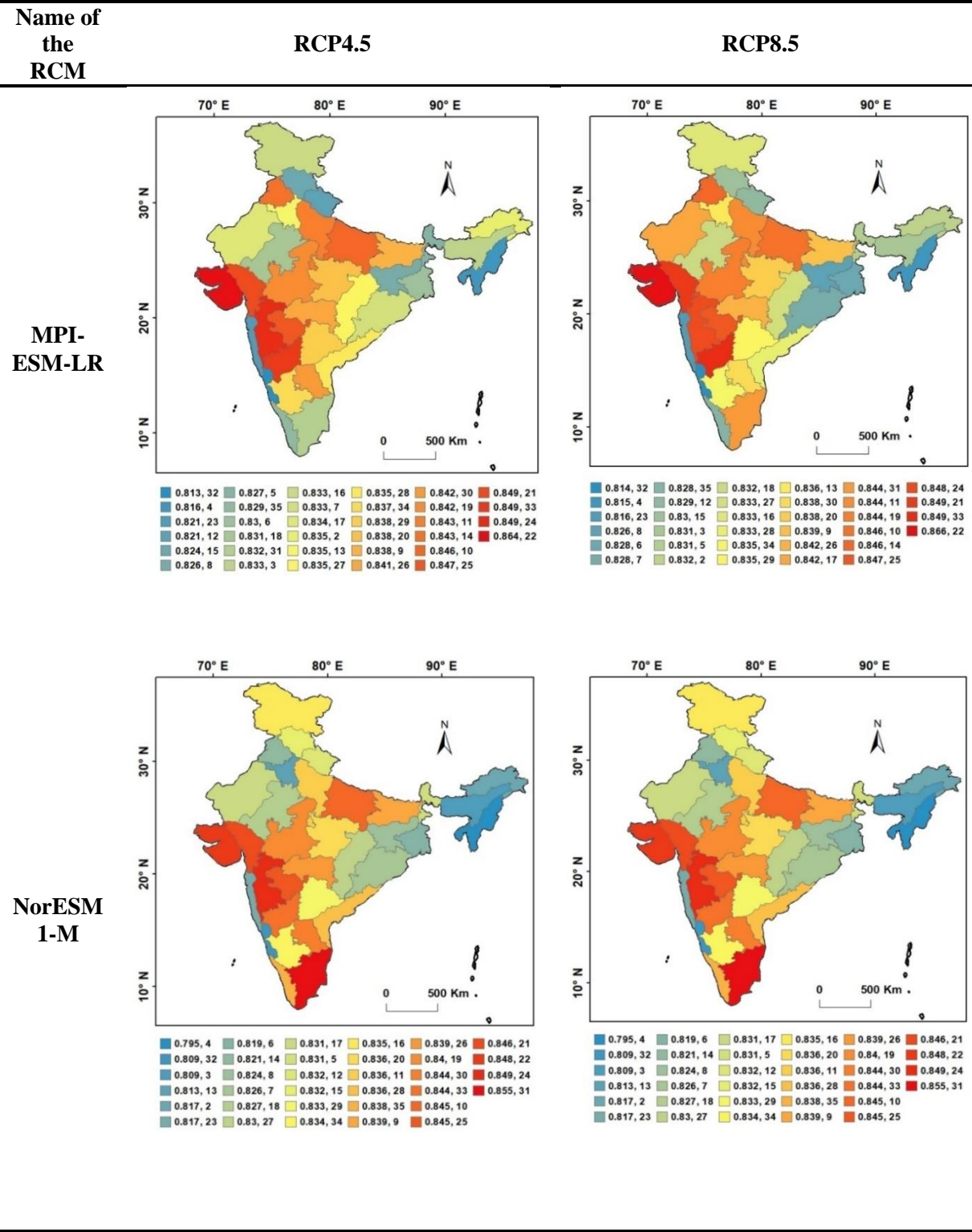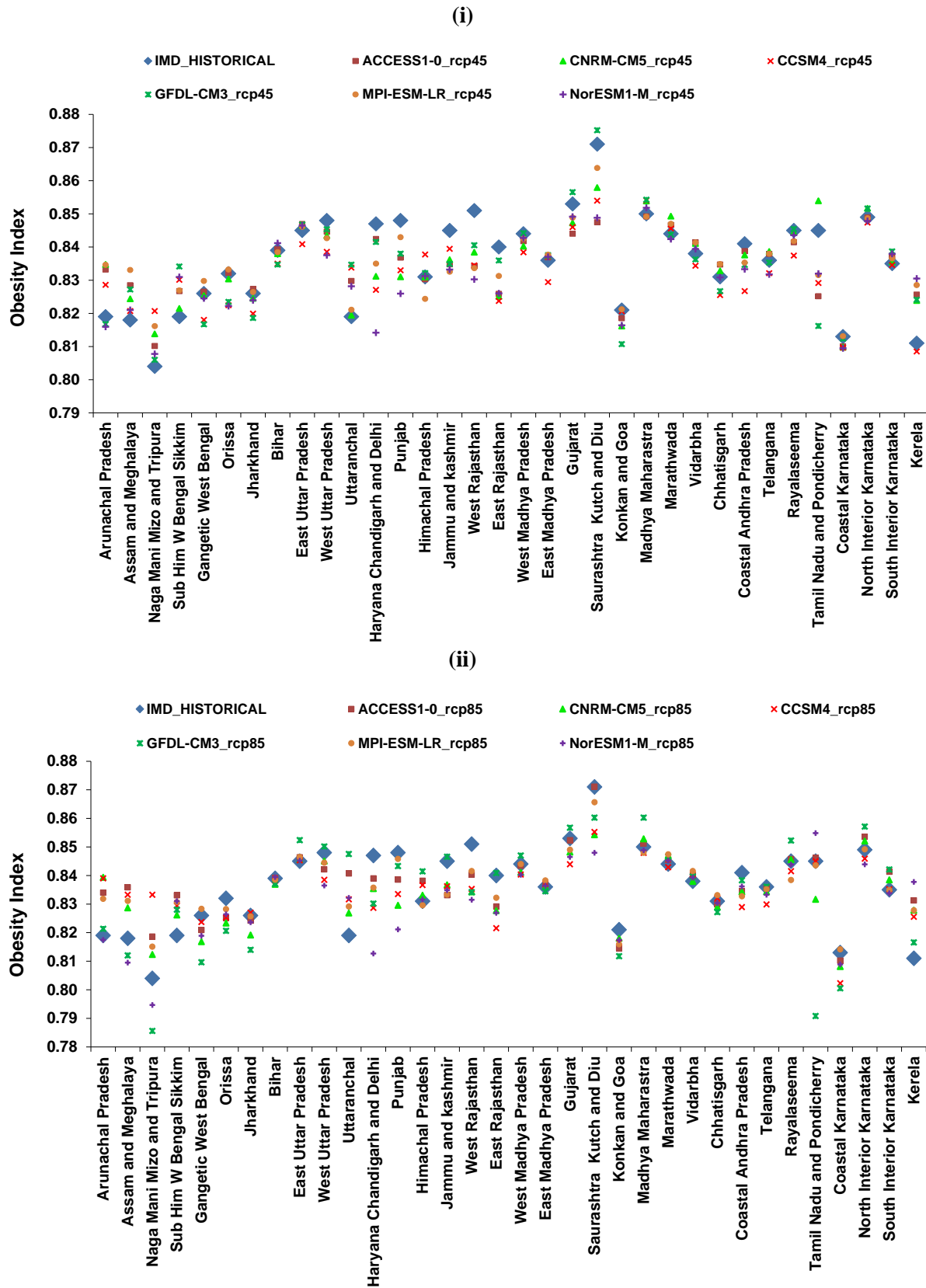
Figure 4.13 Figure showing changes in the mean of obesity index for 34 Meteorological Subdivisions from historical to future time periods for two different scenarios (i) RCP4.5 and (ii) RCP8.5

For each subdivision, an ensemble of OB estimates at individual grids was determined by taking the average of OB estimates obtained from 6 RCMs at those grids. Subsequently, the ensemble means

of OB obtained from 6 RCMs for each subdivision was tested for the significant change relative to the mean of OB corresponding to the historical time period based on the Z-test at a 5% significance level. Table 4.4 provides the Z-statistic along with the result from hypothesis testing based on the Z-test.

Table 4.4 Statistics of Z-test and KS-test and the result from hypothesis testing based on both tests. Here, h=1 means to reject the null hypothesis, and h=0 means to accept the null hypothesis

	Z-test				KS-test			
Subdivision	RCP4.5		RCP8.5		RCP4.5		RCP85	
	Z_stat	h	Z_stat	h	KS stat	h	KS stat	h
Arunachal Pradesh	-3.5263	1	-4.4750	1	0.2329	1	0.2603	1
Assam and Meghalaya	-3.6269	1	-3.0338	1	0.2229	1	0.1988	1
Naga Mani Mizo and Tripura	-1.3519	0	-1.0710	0	0.1321	0	0.0943	0
Sub Him W Bengal Sikkim	-2.4206	1	-2.7413	1	0.2833	1	0.2667	1
Gangetic West Bengal	1.6466	0	3.6265	1	0.1613	0	0.2581	0
Orissa	3.7578	1	5.3878	1	0.1944	1	0.2454	1
Jharkhand	1.7140	0	2.3807	1	0.1667	0	0.2018	1
Bihar	0.5030	0	0.1945	0	0.1083	0	0.0828	0
East Uttar Pradesh	-0.0946	0	-1.1739	0	0.0633	0	0.1041	0
West Uttar Pradesh	4.2235	1	3.6571	1	0.2013	1	0.1883	1
Uttaranchal	-2.8274	1	-5.2861	1	0.2169	1	0.3735	1
Haryana Chandigarh and Delhi	5.7317	1	6.0772	1	0.4143	1	0.4714	1
Punjab	5.0250	1	4.8350	1	0.2857	1	0.3187	1
Himachal Pradesh	0.0909	0	-1.5474	0	0.1136	0	0.2045	1
Jammu and kashmir	6.0969	1	5.1122	1	0.2190	1	0.1776	1
West Rajasthan	11.8152	1	11.1615	1	0.3438	1	0.3312	1
East Rajasthan	8.6378	1	7.4330	1	0.2995	1	0.2609	1
West Madhya Pradesh	2.2041	1	1.3571	0	0.1215	1	0.0972	1
East Madhya Pradesh	-0.3142	0	-0.6276	0	0.0714	0	0.0765	0
Gujarat	2.5983	1	2.0778	1	0.1613	0	0.1129	0
Saurashtra Kutch and Diu	8.1894	1	7.7499	1	0.3661	1	0.3552	1
Konkan and Goa	1.0644	0	1.4681	0	0.1429	0	0.1857	0
Madhya Maharastra	-1.1112	0	-0.7237	1	0.1250	0	0.1250	0
Marathwada	-1.0674	0	-1.0320	0	0.2135	1	0.1910	1
Vidarbha	-0.7160	0	-1.2361	0	0.0876	0	0.1314	0
Chhatisgarh	0.2790	0	0.7284	0	0.0444	0	0.0611	0
Coastal Andhra Pradesh	3.0396	1	3.0843	1	0.1680	1	0.2080	1
Telangana	0.2321	0	1.3003	0	0.0600	0	0.1200	0
Rayalaseema	1.5390	0	0.4194	0	0.1556	0	0.0778	0
Tamil Nadu and Pondicherry	4.8499	1	3.0548	1	0.2343	1	0.1943	1
Coastal Karnataka	0.7948	0	1.7522	0	0.2308	0	0.3077	0
North Interior Karnataka	-0.3437	0	-0.7365	0	0.0865	0	0.1346	0
South Interior Karnataka	-0.6755	0	-1.0339	0	0.1057	0	0.1220	0
Kerela	-3.4196	1	-4.6236	1	0.3175	1	0.3492	1

Also, a KS test at a 5% significance level was performed to verify the considerable change in the distribution of the ensemble of OB estimates obtained from 6 RCMs with respect to the OB estimates corresponding to the historical time period.

For brevity, Figure 4.14 shows the CDF plots of the ensemble of OB estimates obtained from 6 RCMs, and the CDF of OB estimates corresponding to the historical time period for two subdivisions Arunachal Pradesh and Haryana Chandigarh and Delhi. The plots for other subdivisions can be found in Figure S7 in the supplementary material of Gupta and Chavan (2022). Table 4.4 provides the KS statistic along with the result from the KS test. Based on the Z-test and KS test, the following inferences are drawn.

Meteorological Region	RCP4.5	RCP8.5
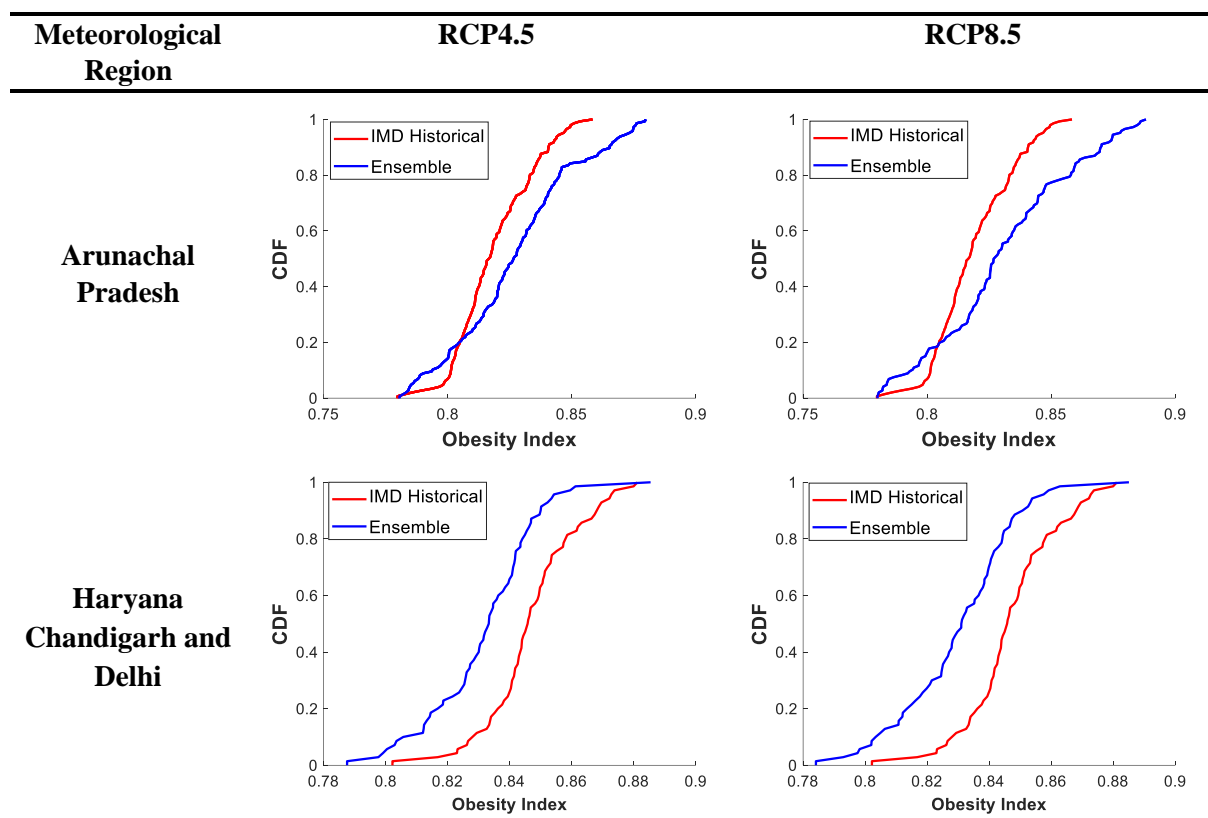
Figure 4.14 Plots showing the CDF of the ensemble of OB estimates obtained from 6 RCMs and CDF of OB estimates corresponding to the historical time period for two subdivisions, namely Arunachal Pradesh and Haryana Chandigarh and Delhi

For both the climate scenarios, the eastern part of India (i.e., Arunachal Pradesh, Assam and Meghalaya, Sub Himalayan, West Bengal, and Sikkim); northern part (i.e., Uttaranchal) and Kerala in South India showed the change in the distribution of the ensemble of OB estimates with reference to the historical time period as well as an increase in the mean of OB for the future time period. Subdivisions like Orissa, West Uttar Pradesh, Haryana, Chandigarh, Delhi, Punjab, Jammu and Kashmir, Rajasthan, Saurashtra, Kutch, Diu, Gujarat, Tamil Nadu and Pondicherry, Coastal Andhra Pradesh showed a change in the distribution of the ensemble of OB estimates with respect to the historical time period and a decrease in the ensemble mean of OB.

The change in the means and CDFs of OB estimates for historical and future time periods were found to be insignificant for some regions like Naga Mani Mizo and Tripura, Bihar, East Uttar Pradesh, Himachal Pradesh, East Vidarbha, Chhattisgarh, Rayalseema, Telangana, Karnataka (Coastal, North and south). Regions like Madhya Maharashtra did not show a significant change in the distribution of ensemble OB for both RCP scenarios; however, it showed an increase in the mean of OB for the RCP8.5 scenario. Jharkhand showed a change in the distribution and mean of OB for RCP8.5 scenario only. Earlier studies (e.g., Maity et al., 2016; Yaduvanshi et al., 2020; Sannan et al., 2020; Todmal et al., 2021; Suman and Maity, 2020; Gupta et al., 2021) indicating an increase/decrease/no change in the frequency of occurrence of precipitation events under a different climate scenario in some of the regions mentioned above supports the finding of this study. Yaduvanshi et al. (2020) identified tropical, temperate, and semi-arid regions of India as the regions showing a significant increase in heavy rainfall-related extreme indices of the Expert Team on Climate Change Detection and Indices (ETCCDI) under two scenarios, RCP4.5 and RCP8.5. Maity et al. (2016) also reported an increase in heavy rainfall in North-Eastern regions of India (such as Assam, Mizoram, Nagaland) under the RCP4.5 scenario. Sannan et al. (2020) found an enormous decrease in Northeast monsoon rainfall over the South Peninsular region (especially Costal Karnataka) projected by most of the CORDEX (CSIRO-CCAM) models in his study. However, some models used in the study showed an increase in heavy rainfall over the southern part of rainfall for both the climate scenario RCP4.5 and RCP8.5. Rai et al. (2020) found that the heavy precipitation indices over India tend to increase more frequently than mean precipitation indices and showed an increase in 1-day maximum precipitation and daily intensity index over some sub-divisions of the west coast, Hilly, and Northeast region.

4.4.6 *Spatial variation of OB and relation with elevation*

Papalexiou et al. (2018) studied the spatial variability of the Pareto and Weibull tail indices with elevation over the US. They reported a coherent pattern with mountainous areas exhibiting heavier tails in high elevations. However, they also noted several other interacting factors such as synoptic conditions, vertical stability, and moisture convergence, which might affect tail heaviness variation. They arranged the tail indices as per the ascending elevation order and considered consecutive blocks of n-points (where n varied from 10 to 400 with an increment of 10). Then the mean elevation (H) and mean tail indices were computed for each block. They observed a convex decrease and concave increase for Pareto and Weibull tail indices, respectively, relative to H. In the context of India, the extreme precipitation usually varies with topography, i.e., elevation (Chavan and Srinivas, 2015). Hence, a similar analysis was performed to explore the spatial variation of OB with respect to the elevation. Plots depicting the clouds of points related to different block sizes (i.e., n varied from 10-400) indicated a nonlinear decrease till a mean elevation of 2500 and showed a non-linear increase beyond that for most of the block sizes. For brevity plots corresponding to four block sizes i.e., 30, 90, 150, and 300 are shown in Figure 4.15, while plots corresponding to other block sizes can be found in Figure S8 of the supplementary material of Gupta and Chavan (2022)).

Table 4.5 presents the ranges of the Obesity Index and elevation (meters) of the grids within each subdivision. The elevation beyond 2500 m is usually found in the subdivision of Jammu and Kashmir, Himachal Pradesh, Arunachal Pradesh, Sub-Himalayan regions, which belong to the Category E (i.e., polar or snow region) according to the Köppen-Geiger classification system. Köppen-Geiger classification system considers meteorological variables like temperature and precipitation to define the climate class of a region. The Köppen-Geiger climate classification system over India consists of four main climate groups: A (tropical), B (dry), C (temperate), and E (polar). The variability in the OB values was observed to be sensitive to the climatic conditions or regional physical processes. Papalexiou et al. (2018) observed the heaviest tails for mountainous regions (belonging to the category of snow), which supports the findings of this study also. However, for lower elevation than 2500 m, we did not find similar results as Papalexiou et al. (2018) because the regions for which elevation is less than 2500 m comprises regions belonging to arid, temperate, and tropical zones in India (Yaduvanshi et al., 2021). The heaviest tails were observed in Arid and Semi-Arid (e.g., Saurashtra, Kutch and Diu, Gujarat, West Rajasthan, Punjab, Haryana Chandigarh and Delhi, Madhya Maharastra, parts of North Interior Karnataka) which have different elevations ($H < 2500$). Relatively heavier tails were found in Tropical regions comprising Southern and Southern Eastern states of India (e.g., Tamil Nadu and Pondicherry, Coastal Andhra Pradesh, parts of Rayalseema, Marathwada). Moreover, slightly lighter tails were observed in Temperate regions (e.g., Naga Manipur, Mizoram and Tripura, Assam and Meghalaya, Jharkhand) and Tropical regions comprising parts of Western Ghats (Konkan and Goa, Coastal Karnataka, Kerela) of India. Papalexiou et al. (2018) also found relatively thin tails in the coastal plains, just like our finding of thin tails in coastal areas (west coastal area). The difference in the heaviness of the tails between the Eastern coast and Western coast (both belonging to Tropical regions) might be attributed to many interacting factors such as synoptic conditions, moisture convergence, topography (see Figure S9 in the supplementary material of Gupta and Chavan, (2022)), different extreme rainfall characteristics such as mean annual maximum precipitation (see Figure S10 in the supplementary material of Gupta and Chavan, (2022)), and seasonality (see Figure S11 in the supplementary material of Gupta and Chavan, (2022)) (which represents the day of occurrence of extreme precipitation in a year) despite having similar elevation range. Interestingly, Suman and Maity (2020) also observed the differences in mean extreme precipitation on the eastern coast of south India than the western coast for the observed record. It is noteworthy that along with the relationship of OB with elevation, we found some spatial coherence of OB with the climate classes of the Köppen-Geiger classification system.
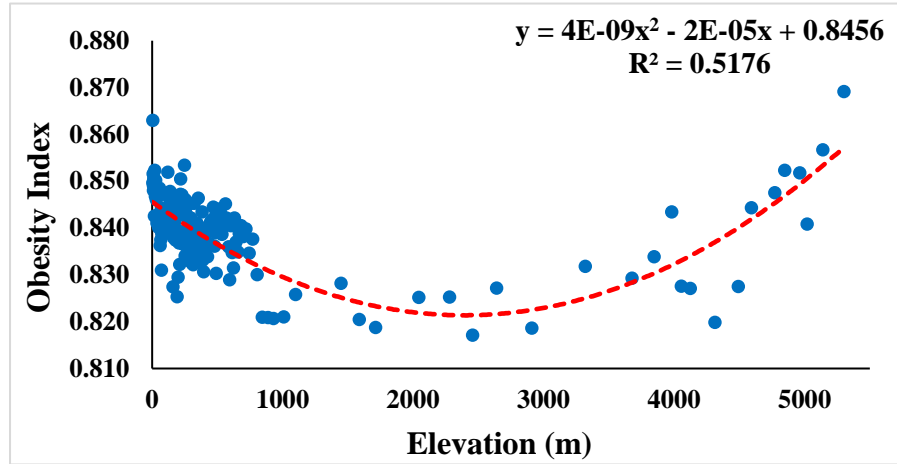
Figure 4.15 Plots showing the variation of Obesity Index *vs*. Elevation (*H*) for four different block sizes (i) Block size *n=30*, (ii) Block size *n=90*, (iii) Block size *n=150* and (iv) Block size *n=300*

Table 4.5 Information on obesity index and mean elevation (*m*) for the grids within each Meteorological Subdivision

ID	Meteorological Region	Obesity Index		Elevation	
		Minimum	Maximum	Minimum	Maximum
2	Arunachal Pradesh	0.779	0.858	385.32	4830.56
3	Assam and Meghalaya	0.774	0.877	35.38	1656.50
4	Naga Mani Mizo and Tripura	0.767	0.851	13.29	1237.83
5	Sub Him W Bengal Sikkim	0.772	0.859	23.72	4993.52
6	Gangetic West Bengal	0.803	0.860	2.73	313.88
7	Orissa	0.799	0.870	3.12	984.37
8	Jharkhand	0.789	0.861	51.64	833.95
9	Bihar	0.804	0.870	35.38	810.67
10	East Uttar Pradesh	0.792	0.878	62.62	810.67
11	West Uttar Pradesh	0.811	0.871	112.48	476.00
12	Uttaranchal	0.785	0.861	304.26	5096.27
13	Haryana Chandigarh and Delhi	0.802	0.881	185.93	476.00
14	Punjab	0.811	0.899	171.12	2084.49
15	Himachal Pradesh	0.798	0.872	320.95	4918.10
16	Jammu and kashmir	0.778	0.930	454.23	5594.08
17	West Rajasthan	0.811	0.893	19.01	448.24
18	East Rajasthan	0.806	0.881	134.84	803.48
19	West Madhya Pradesh	0.796	0.883	131.51	794.98
20	East Madhya Pradesh	0.793	0.867	157.23	842.00
21	Gujarat	0.818	0.883	2.81	487.27
22	Saurashtra Kutch and Diu	0.827	0.906	0.21	279.96
23	Konkan and Goa	0.782	0.864	8.02	751.87
24	Madhya Maharastra	0.809	0.881	148.36	920.78
25	Marathwada	0.819	0.869	339.86	712.00
26	Vidarbha	0.814	0.866	131.32	809.93
27	Chhatisgarh	0.795	0.868	133.96	900.88
28	Coastal Andhra Pradesh	0.788	0.885	3.87	1035.36
29	Telangana	0.799	0.862	77.76	652.08
30	Rayalaseema	0.804	0.876	39.64	793.40
31	Tamil Nadu and Pondicherry	0.784	0.892	2.36	4877.20
32	Coastal Karnataka	0.783	0.842	8.00	876.67
33	North Interior Karnataka	0.819	0.880	351.48	756.61
34	South Interior Karnataka	0.778	0.885	68.71	1059.13
35	Kerela	0.776	0.858	4.23	3975.00

4.5   *Summary and Conclusions*

In this chapter, an algorithmic approach based on the concept of the OB was used to characterize the tails of probability distributions of 0.25° daily gridded precipitation data over India. We performed a simulation study that recommended that the optimum sample size greater than or equal to 1000 and a minimum number of random sampling equal to 5000 are needed to determine a consistent OB estimate. A comparative analysis between the OB-based approach and recent threshold-based approaches

proposed by Nerantzaki and Papalexiou (2019) and Papalexiou et al. (2013) was conducted. All three approaches provided consistent inferences regarding heavy-tailed behavior over India. The relative assessment of the results revealed that around 78.21% of grids showed heavy-tailed behavior based on all three approaches. Further, the pairwise assessment revealed that the OB-based approach tends to provide comparatively lighter tails for the grids where threshold-based approaches yielded exponential tails. However, the three approaches might not always provide similar assessments as they inherently differ from each other in the way they analyze the tail behavior of daily precipitation data. Overall, all three approaches showed that heavy-tailed distributions provide a better fit than light-tailed for daily precipitation extremes over India. The major advantage of the OB-based approach lies in its ability to provide a quantitative measure to assess the tail behavior of daily precipitation data without assuming any threshold value to segregate tails.

The application of the OB-based approach was extended to characterize tails of future projections of daily gridded precipitation obtained from CORDEX-SA experiments comprising 6 RCMs for RCP4.5 and RCP8.5 scenarios of climate change. In the case of the RCP8.5 scenario, a higher frequency of occurrences of extreme precipitation was observed as compared to the RCP4.5 scenario. We extended the utility of the OB-based approach to characterize daily precipitation data of meteorological subdivisions over India. The subdivision-wise analysis highlights the necessity to consider heavy-tailed distributions for modelling extreme precipitation events for reliable planning and design of hydrological structures within the subdivisions. This analysis could be useful for the identification of apt regional frequency distributions for the meteorological subdivisions. Finally, we explored the spatial variation of OB with respect to the elevation over India. It was observed that the OB tends to decrease non-linearly with an increase in the elevation till 2500 m and then showed an increase beyond the elevation of 2500 m. The spatial change in OB over elevation showed spatial coherence with the climate classes of the Köppen-Geiger classification system over India.

In summary, the results from this study highlight the importance of considering heavy-tailed distributions instead of the traditional light-tailed distributions for the reliable estimation of the frequency of extreme precipitation events that can find use in the design of major civil engineering infrastructure.

A COMPREHENSIVE DECISION SUPPORT SYSTEM FOR THE CHARACTERIZATION OF PROBABILITY DISTRIBUTION TAILS FOR DAILY PRECIPITATION

5.1   Overview

A practical approach to select the appropriate class of distribution for a dataset considering its right tail (i.e., extreme) is termed a Decision support system (DSS). A DSS utilizes various graphical methods to characterize the tails of distributions as per their limiting behavior and groups them into the appropriate class of distributions. Popular classes of the probability distributions are given by Ouarda et al. (1994) and Werner and Upper (2004) in a nested form as $A \subset B \subset C \subset D \subset E$, starting from light-tailed distributions to heavy-tailed distributions) (as presented in Figure 1.1 in supplementary material). All distributions with tails decaying more slowly than the exponential tail are included in class D. The tails of distributions in class C decay asymptotically according to the power law. Class B contains the distributions having the exact Pareto tail. Lognormal (LN) is a limiting case with a tail between classes D and C (Champernowne, 1953; Martel et al., 2013).

The conventional DSS proposed by El Adlouni et al. (2008) and Ehsanzadeh et al. (2010) utilizes various graphical approaches such as the Mean Excess Function (MEF) plot, log-log plot, and the generalized Hill ratio plot to characterize the tails of distributions. However, existing DSSs lack efficient segregation of the Lognormal distribution from the Regularly varying and Subexponential distribution families. Also, they lack the ability to identify the distributions from the hyper-exponential distributions. Recently developed graphical diagnostic tools, such as concentration profile, concentration adjusted expected shortfall, discriminant moment ratio plot, maximum-to-sum plot, and Zenga plot can classify the tails of distributions into various classes if used in an appropriate order in combination with tools of conventional DSS. This chapter presents a comprehensive DSS that is proposed to alleviate the shortcomings associated with the conventional DSS and characterizes the tails of distributions into classes B\A (Pareto type), C\B (regularly varying), D\C (subexponential), E (Exponential type), hyper-exponential class (outside class E) and LN (Lognormal) distribution (the limiting case between class C and D). The robustness of the proposed DSS over the conventional DSS is established through a simulation experiment. Further, this study also evaluates the influence of the sample size on the effective implementation of the proposed DSS. Finally, the proposed DSS is applied to characterize the tails of daily gridded precipitation data over India

5.2 *Shortcomings of the Conventional DSS*

Conventional DSS partitions the dataset into various classes/families of distributions. They assume that a probability distribution either belongs to a light or a heavy-tailed distribution class, which may not be valid always. Recent studies by Cooke and Nieboer (2011), Weitzke et al. (2020), and Gupta and Chavan (2022) have shown that a particular probability distribution can be characterized as a light or heavy-tailed distribution based on its shape parameter or tail index. For example, the Gamma distribution with shape parameter $\alpha > 1$ and $\alpha < 1$, possess a "slightly lighter" and "slightly heavier" tail relative to the exponential tail, respectively, while it has an exact exponential tail when $\alpha = 1$. Similarly, Cirillo and Taleb (2020) described the erratic behavior of LN distribution and compared it to the analogy of a wolf in sheep's clothing. In fact, the LN distribution possesses a thin tail for the shape parameter $\alpha < 1$ or exhibits a very heavy tail for $\alpha > 1$ (May et al., 2013; Cirillo, 2013; Nerantzaki and Papalexiou, 2019; Cirillo and Taleb, 2020; Gupta and Chavan, 2022). Gupta and Chavan (2022) presented the relative measure of the tail heaviness for precipitation data over India based on a scalar tail index such as the Obesity index. They demonstrated that the Weibull, Gamma, or LN distributions could possess more heavy or obese tails than a Pareto distribution, depending on the choice of shape parameter or tail index.

Moreover, El Adlouni et al. (2008) and Ehsanzadeh et al. (2010) mentioned that LN distribution lies at the frontier of classes C and D. However, they both considered LN as a part of class D due to (i) more conservative results (i.e., overestimation) in the quantile estimation study and (ii) unavailability of suitable criteria to discriminate amongst the classes C, D, and LN (see Figure 5.1). Recently, Martel et al. (2013) inferred that the association of LN to class D could result in significant errors in quantile estimation. They illustrated the significant effect that the choice of distributions from the classes C, D, and LN have on estimated quantiles corresponding to the T-year return period for any hydroclimatic variable. Hence, they improved the conventional DSS by adding a new step to the DSS where the lognormality check was performed before assessing the adequacy of the classes C or D (see Figure 5.2).
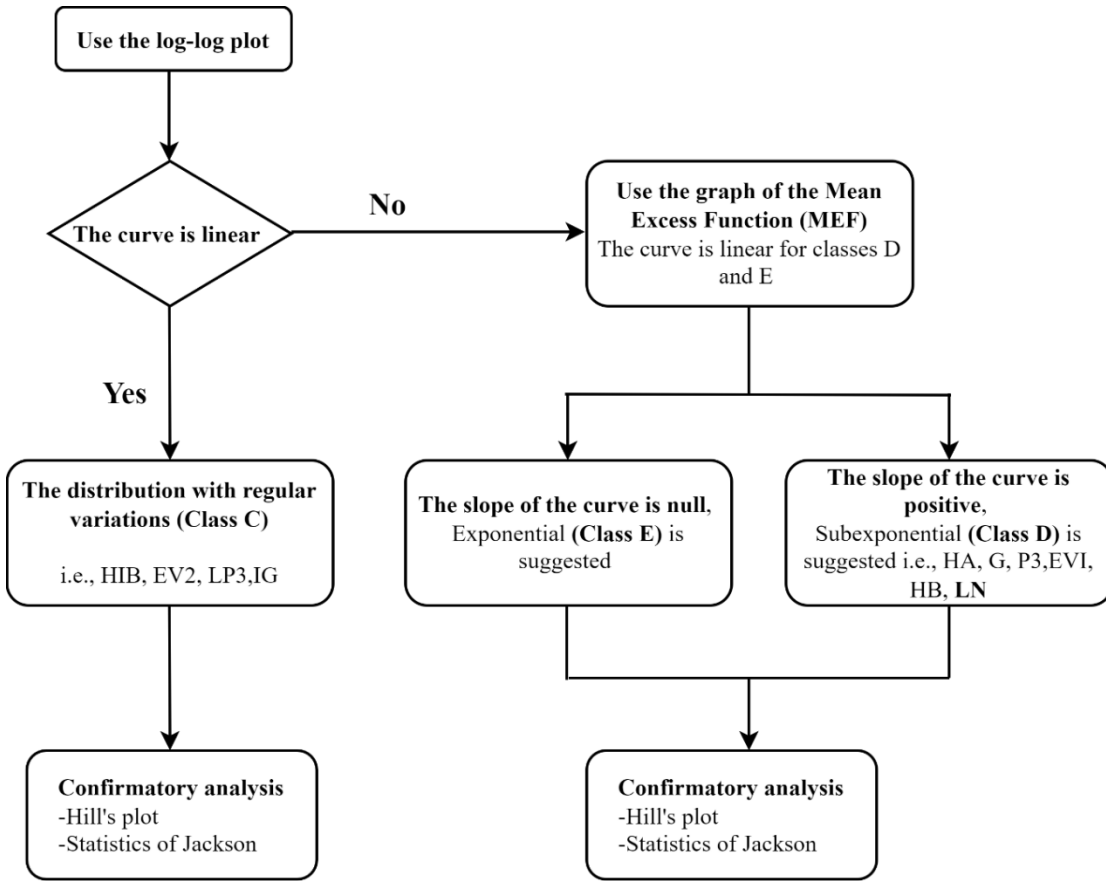
Figure 5.1 Flow diagram of the first version of DSS (see Ehsanzadeh et al., 2010)
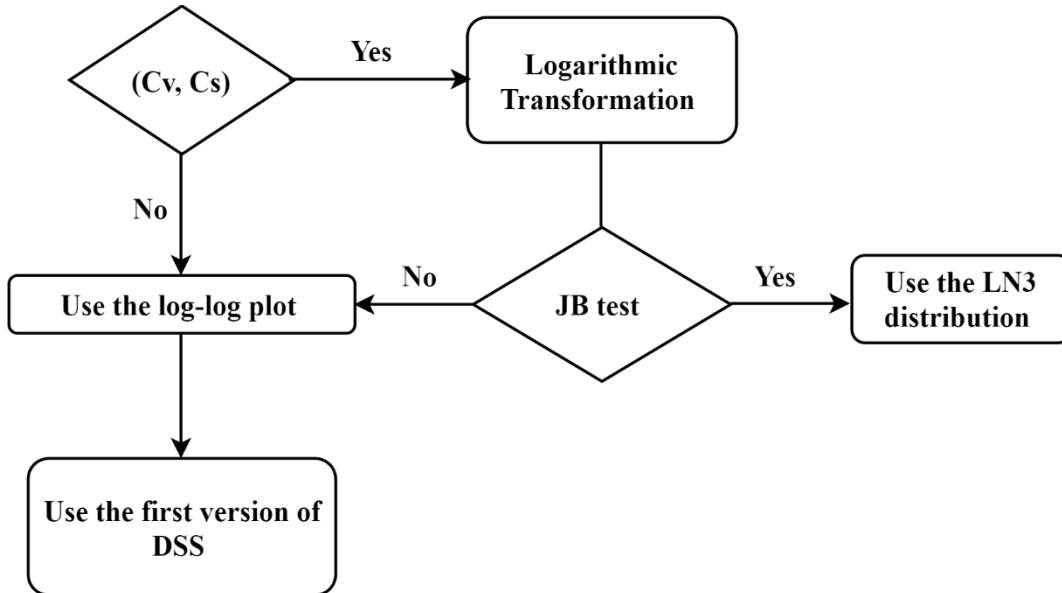
Figure 5.2 Flow diagram of the second version of DSS using two-dimensional (C_v, C_s) plot and Jarque-Bera test (see Martel et al., 2013)

Firstly, they considered a two-dimensional plot between the coefficients of variation (C_v) and skewness (C_s) and checked the position of the sample in the diagram. If the sample falls in the Halphen

Type IB (HIB) area, a subset of the LN zone, then the Jarque-Bera (JB) (a normality test) is applied to the log-transformed data. It was an effective procedure to discriminate the LN3 from the Classes C and D distributions; however, there is subjectivity associated with the use of the (C_v, C_s) plot, being a range-bound plot $(C_v \in (0,1), C_s \in (1, 2.5))$, and the value of C_v and C_s may exist beyond these ranges for the real and large database like daily precipitation. Further, the statistical tests used to test normality have the disadvantage of not being sensitive enough for low sample sizes or overly sensitive to large sample sizes (Mishra et al., 2019). Besides, using transformations, such as the log transformation, can be quite problematic. The researchers need to be mindful of the limitations while analyzing the transformed data instead of the original data (Changyong et al., 2014). In summary, various studies have highlighted the necessity to segregate the LN distribution from the classes C and D (it can possess either light or heavy tails) and characterize the probability distributions of datasets so that reliable estimation of the quantiles of the variables corresponding to various exceedance probabilities can be achieved. There is a need to characterize the tails of empirical datasets effectively into several classes/families through the construction of a comprehensive DSS.

5.3   *Recent Advancements in Graphical methods*

Graphical methods encompass tools that are useful for exploratory analysis and tail discrimination (Nerantzaki and Papalexiou, 2019) and form an essential component of the DSS. Graphical approaches such as MEF plot, log-log plot, and generalized Hill ratio plot were part of the conventional DSSs. However, lately, studies have developed more advanced ways of diagnosing the distribution tails, which could be useful to characterize the distributions into different classes (Zenga, 2007; Nair, 2012; Cirillo, 2013; Fontanari et al., 2018a). Cirillo (2013) introduced the Discriminant moment ratio plot (DMR) and showed that a simpler version of the CV-Skewness diagram (Vargo et al., 2010) is efficient enough to scrutinize the inferences revealed from the log-log and MEF plots. It should be noted that the two-dimensional (C_v, C_s) plot used by Martel et al. (2013) is a subset of the DMR plot, which, unlike the (C_v, C_s) plot, extends over a high range of C_v and C_s. In the plot, points or curves represent some distributions, while the generalized distribution and families of distribution are represented as zones. The entire plot is divided into 4 zones, namely, (i) The Paretian zone, (ii) The Gray zone, (iii) The Lognormal zone, and (iv) The exponential/Thin-tailed zone. The location of the couple point (C_v, C_s) or $(\hat{\gamma}_2, \hat{\gamma}_3)$ with respect to the four areas and curve gives us a good idea of possible candidate distribution. Cirillo (2013) also proposed the Zenga plot, which helps to discriminate between the lognormal and the Pareto distributions. Zenga curve can be expressed analytically for any distribution via the corresponding Lozern curve ($L(u)$) (Lorenz, 1905)). Zenga plot shows a Zenga curve ($Z(u)$) plotted against u (threshold varying between 0 and 1) and assumes different shapes for different distributions. For Pareto distribution $Z(u)$ is a convex increasing curve on [0,1] and approaches the u-

axis for larger shape parameters, indicating a decrease in concentration. While for an LN distribution $Z(u)$ remains constant and changes with the variance (σ).

Another interesting graphical tool to perform tail discrimination is the maximum-to-sum plot (MS plot or $R_n(p)$) which is a ratio of partial maximum (M_n^p) and partial sum (S_n^p) for p statistical moments (Herein $p=1$: mean; $p=2$: variance; $p=3$: skewness; and $p=4$: kurtosis). The power-law classes (Class B and Class C) are characterized by the non-existence of higher-order moments, while distributions belonging to LN and class D\C have the existence of all their moments. The MS Plot relies on a simple consequence of the law of large numbers (Embrechet et al., 2003; Cirillio and Taleb, 2020; Manz and Mansmann, 2020) and sees the non-convergence or convergence of $R_n(p)$ to zero for inferences about power-tail or non-power tail type distributions, respectively.

Subsequently, Fontanari et al. (2018a) introduced concentration profile (CP) plots along with Concentration adjusted expected shortfall (CAES) plots, which were used to identify the parametric families of the loss distribution, especially Lognormal, Pareto, Exponential, and Weibull ($\alpha > 1$ and $\alpha < 1$) behavior. These plots were based on concentration (or inequality) measures (e.g., the Gini index derived from the Lorenz curve and common risk management measures like the Expected Shortfall (ES) (Acerbi and Tasche, 2002; McNeil et al., 2015; Cirillo and Taleb, 2016). Gini index has been applied in the field of hydro-meteorology to capture the inequality and temporal changes in distributions of precipitation (Rajah et al., 2014; Monjo and Martin-Vide, 2016; Konapala et al., 2017; Shrestha et al., 2019), streamflow (Masaki et al., 2014; Zhang et al., 2015), and river solute loads (Jawitz and Mitchell, 2011; Preisendanz et al., 2020) at a different scale. Recently, the Gini index has been proposed as a reliable upper tail indicator to characterize the heavy tail phenomena (Fontanari et al., 2018b; Wietzke, 2020). However, Fontanari et al. (2018a) preferred the sequence of truncated Gini indices (indicated by $G(\lambda)$, where λ denotes truncation level) over the conventional Gini index, as it measures the dispersion above the Value at Risk (VaR) (Jorion, 2001) to have a reliable measure of tail risk and precision of the ES.

All new graphical methods are robust in assessing the tail behavior of the distributions. However, there is a dearth of attempts to incorporate these advanced graphical methods into the DSS to improve the characterization of tails of datasets. Hence, this study is envisaged to devise a comprehensive DSS that would incorporate all the advanced graphical methods to characterize the tails of empirical datasets into various classes. If put to use in the correct order, these graphical methods can be very useful in identifying the true nature and behavior of the datasets and characterizing the tails of various classes, especially when the competing tails are present to choose from, e.g., lognormal and regularly varying tails. This makes the use of various graphical methods very appealing for formulating a comprehensive DSS. Details on the all aforementioned graphical methods used in Conventional DSS along with the one proposed to be used in comprehensive DSS can be found in the next section 5.4 for better understanding.

5.4 Details of the Graphical methods used in DSS to characterize tails

Detailed investigation of various graphical methods incorporated in the comprehensive DSS along with their properties can be found in Embrechts et al. (2003), Beirlant et al. (2006), El Adlouni et al. (2008), Ehsanzadeh et al. (2010), Cirillo et al. (2013), Nerantzaki and Papalexiou (2019) and Fontanari et al. (2018a). A brief description of the methods is presented in the following sections.

5.4.1 Mean Excess Function (MEF)

The plot of the mean excess function (Embrechts et al., 2003) is based on the behavior of a function $e(u)$ given as:

$$e(u) = E[X - u \mid X > u] = \frac{\int_u^\infty (t-u)dF(t)}{dF(t)}, \qquad 0 < u < x_F \tag{5.1}$$

where X is a random variable with distribution function F and right endpoint x_F such that $x_F = \sup\{x \in \mathrm{R} : F(x) < 1\}$. The function $e(u)$ is called the mean excess function of X. From an empirical point of view, the MEF of a sample $X_1, X_2, \ldots, X_n$ is easily computed as

$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)}{\sum_{i=1}^{n} 1_{\{X_i > u\}}} \tag{5.2}$$

Equation (5.2) represents the ratio between the sum and the number of exceedances over the threshold u. If the empirical value of $e_n(u)$ $vs.$ u is linear and the slope is equal to zero or null, it indicates an exponential tail. If the plot shows a slope greater than zero, it might indicate subexponential tails, as the mean excess function tends to infinity for such tails (Cooke and Nieboer, 2011). If the plot shows a slope lesser than zero (i.e., negative), the tail is of hyper-exponential type (Nerantzaki and Papalexiou, 2019). Hence, in this study, this plot discriminates the distributions of class D with that of class E\D and also identifies hyper-exponential tails (i.e., tails outside of class E). To estimate the MEF slope the MEF slope the advancements suggested by Nerantzaki and Papalexiou (2019) was implemented. Nerantzaki and Papalexiou (2019) considered two statistical tests in their study that are followed in this study also. Firstly, a two-tailed test is performed to test the null hypothesis that the tail of daily precipitation at a station is exponential. The test utilizes a confidence interval (CI) for the exponential case corresponding to sample size (n) for a specific significance level (u). If the estimated slope lies within the CI, then the null hypothesis cannot be rejected for the selected significance level. Contrary to this, if the estimated slope lies outside the CI, then the null hypothesis is rejected, indicating a non-exponential tail. When the null hypothesis from the two-tailed test is rejected, a one-tailed test is performed where the null hypothesis is Exponential tail, and the alternative is sub-exponential (or hyper-exponential). Tails heavier than the exponential tail, i.e., the observed slope is above the CI's upper limit

are designated as the sub-exponential tails, whereas the tails lighter than the exponential tail, i.e., the observed slope is below the CI's upper limit are designated as the hyper-exponential tails. Finer details about the approach can be found in Nerantzaki and Papalexiou (2019).

5.4.2 *Concentration Profile and Concentrated Adjusted Expected Shortfall*

A CP plot is a tool based on the concept of using the concentration (or inequality) measure to analyze the risk dispersed in the tail. A CP is nothing but a sequence of truncated Gini indices that could be used to discriminate among different tail risk profiles and characterize different probability distributions. The truncated Gini index is defined as

$$
\begin{aligned}
G(\lambda) &= 1 - 2\int_0^1 L_\lambda(x)dx \\
&= 1 - 2\int_0^1 \frac{L(\lambda + (1-\lambda)x) - L(\lambda)}{1 - L(\lambda)}dx
\end{aligned}
\tag{5.3}
$$

Where, $L_\lambda(x)$ is defined as the truncated Lorenz curve and λ is the truncation level (or confidence level). Also, the Gini index can be defined in terms of Value at Risk (VaR_λ) and Expected Shortfall (ES_λ) for a positive Y with CDF $F(y)$, Equation (5.3) can be re-written as

$$
G(\lambda) = \frac{E(|Y_1 - Y_2| \, | \, Y_1 \wedge Y_2 > VaR_\lambda)}{2ES_\lambda} \in [0,1]
\tag{5.4}
$$

As a consequence of this procedure, for a fixed λ, $G(\lambda)$ corresponds to the Gini index of a new random variable X with support $[VaR_\lambda \ c)$, with $c \leq \infty$, and expectation ES_λ. This truncated Gini index measures the dispersion above VaR_λ for a fixed confidence level λ such that a reliable measure of tail risk can be defined along with ES_λ precision. The truncated Gini index $G(\lambda)$ inherits all the properties of the usual Gini index and is, therefore, a measure of tail dispersion. However, since the main interest is usually the right tail where the large value (i.e. extreme) lies, the above formulation of truncated Gini is much better than the usual Gini index as it takes into account the support of the distribution based on the truncation level rather than including the support of entire distribution. A high value of $G(\lambda)$, i.e., closer to 1, implies that the losses in the risk subclass (S_λ) are dispersed, resulting in a persistent tail thickness in the subclass and a lower precision of the corresponding ES_λ. A value close to 0 suggests that losses are less dispersed within the risk subclass. Graphically, CP (denoted as $\{G(\lambda)\}_{\lambda \in [0,1]}$) is obtained by plotting $G(\lambda)$ against the increasing value of λ.

To form a CP for a sample (with sample length *n*), all the observations in a sample are sorted in increasing order, and the Gini index is computed *n* times recursively, each time excluding the first *n* smallest observation. For an accurate estimation using this method, a sufficient number of observations

is left in the right tail, denoted by k; generally, k varies between 1-5% of the originally ordered data points. Graphically, CP (denoted as $\{G(\lambda)\}_{\lambda \in [0,1]}$) is obtained by plotting $G(\lambda)$ against the increasing value of λ. CP can very well characterize or identify distribution belonging to different classes and provide a rationale for their discrimination based on its shape. It allows the characterization of class B\A (i.e., a Pareto-type tail) and can also identify distribution from class E, D\C, and the limiting distribution like lognormal (LN). The theoretical CP for some distributions from these classes is given in Fontanari et al. (2018a).

An interesting observation obtained from the CPs for different tail types is that for the fat or heavy tail domain, the risk does not vary much like for purely fat-tailed distribution. The Pareto distribution is characterized by a constant CP, where the height of the line only depends on the shape parameter of the distribution itself. Paretian losses maintain their inherent level of risk, which is inherently higher when the tail is heavier. While for the other distributions like Lognormal, Weibull, and Exponential, the CP exhibits a decreasing behavior (i.e., a non-increasing CP always) indicating that losses above high values of VaR_λ tend to be less dispersed. Exponential CP starts with $G(0) = 0.5$ at truncation level $\lambda = 0$ and decreases towards zero first convexly and successively concavely, with a point of flex $\lambda = 0.63$. A Weibull distribution with a shape parameter $\alpha > 1$ has a CP below the exponential CP, without intersection, while for $\alpha < 1$, the Weibull CP lies uniquely above it. For the lognormal CP, if $\alpha < 1$, the starting point of the lognormal CP is $G(0) < 0.5$; when $\alpha > 1$, we have $G(0) > 0.5$. Compared to the Weibull CP, the lognormal CP is flatter, indicating that its tail decreases more slowly than the latter. For the light-tailed domain, CP exhibits a quick decreasing behavior (quick variability) with $G(0) < 0.5$. All these heuristics makes CP an efficient goodness-of-fit tool that could be easily applied to any set of data for making quick as well as in-depth statistical analysis of the dataset.

Concentration Adjusted Expected Shortfall (CAES) is the product of Expected Shortfall at a confidence level, λ i.e., ES_λ and the corresponding truncated Gini index $G(\lambda)$. CAES shows different behavior for different distributions. CAES tends to infinity for Pareto distribution for any value of shape or tail parameter α, Weibull distribution with $\alpha < 1$, and lognormal distribution with $\alpha \geq 0.3$. CAES shows a constant ($c \in R$) profile for the exponential distribution while the profile tends to zero for Weibull $\alpha \geq 0.3$ and lognormal $\alpha < 0.3$ as the truncation level $\lambda \to 1$. For Lognormal distribution, CAES shows three different behaviors based on the shape parameter α, giving us some idea about the value of the shape parameter of the distribution. For more detailed plots of CP and CAES for different distributions, refer to Fontanari et al. (2018a).

5.4.3 Zenga Curve

Lorenz curve and Gini index are popular measures of inequality used since the nineteenth century (Lorenz, 1905; Gini, 2014; Manz and Mansmann, 2020). As a measure of concentration, the Zenga

curve is an alternative measure to the well-known Lorenz curve. Zenga (2007) related the Zenga curve to the Lorenz curve, as shown in Equation (5.5)

$$Z(u) = \frac{u - L(u)}{u[1 - L(u)]} , \qquad 0 < u < 1 \tag{5.5}$$

Where $L(u)$ is the Lorenz curve for the distribution above a pre-specified threshold defined as

$$L(u) = \frac{1}{\mu} \int_0^u Q(s)ds, \quad u \in [0,1] \tag{5.6}$$

Equation (5.6) presents an analytical form of the Zenga curve. Differently from the Lorenz curve, the Zenga assumes different shapes for different distributions. For Pareto distribution, the Zenga curve is positively-sloped (convex increasing function) and rises as $u \to 1$; the higher the curve, the more heavy-tailed the distribution becomes. For Lognormal distribution, the Zenga curve is a constant horizontal curve depending on the shape parameter α of the distribution (see Equation (5.7))

$$Z(u) = 1 - e^{-\alpha^2} , \quad 0 < u < 1 \tag{5.7}$$

Since Pareto always shows an increasing curve and lognormal a constant curve, the Zenga plot forms a good basis to discriminate between these two distribution (Cirillo, 2013). Hence, in this study, this plot is used to identify the distribution from class B\A and LN. Codes used to produce the Zenga curve are referred from Cirillo (2013). Note that some curvatures at extremities may appear in the Zenga curve for the lognormal distribution, depending on the empirical computation of the curve, and become less and less relevant as the number of observations increases.

5.4.4   *Discriminant moment ratio plot (DMR plot)*

The moment ratio plot, introduced by Craig (1936) and further developed by Johnson and Kotz (1970) and Vargo et al. (2010), provides a simple way of visualizing distributions and discriminating among them. Cirillo (2013) showed that the best moment ratio plot is a simpler version of the CV-Skewness moment ratio plot in which the information related to a given distribution is summarized by the behavior of the couple point (C_v, C_s) or $(\hat{\gamma}_2, \hat{\gamma}_3)$. In the plot, some distributions are represented as points, while others are shown as curves and in cases of generalized distribution and families of distribution as areas. The plot is divided into 4 areas, namely, (i) The Paretian zone, (ii) The Gray zone, (iii) The Lognormal zone, and (iv) The exponential/Thin-tailed zone. The equations of the couples representing different curves, thereby defining the zones, are referred from Cirillo (2013). The location of the couple point (C_v, C_s) or $(\hat{\gamma}_2, \hat{\gamma}_3)$ with respect to the four areas and curve gives us a good idea of possible candidate distribution. If the point falls in the Paretian or lognormal zones (close to the lognormal curve), it is likely to be Pareto and Lognormal types, respectively. If the points fall in the

92

Exponential/Thin Tailed zone, both the lognormal and the Pareto are completely ruled out. More analyses are needed in cases where the point falls in the so-called Gray zone, as this area concerns mixtures of lognormal and power tail and lognormal with extremely large variances. A point falling out of the four areas may represent a symmetrical or normal distribution if it lies close to the dotted curve or a mixture thin-tailed distribution if it falls below the Bernoulli curve. Furthermore, the benefit of the discriminant moment-ratio plot is that it allows the simultaneous comparison of many different distributions. In this study, a DMR plot is used to identify the LN distribution distinctively along with other distributions from class C\B, E, and Hyper-exponential family (outside class E).

5.4.5 The maximum-to-sum plot (MS plot)

The "maximum-to-sum" (*MS*) plot is based on a simple consequence of the law of large numbers (Cirillio and Taleb, 2020; Manz and Mansmann, 2020). For a sequence $X_1, X_2, ..., X_n$ of non-negative independent and identically distributed random variables, the maximum-to-sum ratio is given as

$$R_n^p = M_n^p / S_n^p \to 0 \qquad\qquad , n \geq 1, \quad p > 0, \qquad\qquad (5.8)$$

Where, $M_n^p = \max(X_1^p, ..., X_n^p)$ is the partial maximum of order p and $S_n^p = \sum_{i=1}^{n} X_i^p$ is the partial sum of order p. The moment of order p of the distribution exists, i.e., $E[X^p] < \infty$ for $p = 1, 2, 3.....$, if and only if $R_n(p)$ converges to zero for $n \to \infty$. In practice, we plot R_n against n for various values of p. If the $R_n(p)$ jumps up and does not converge to zero for any p exceeding p_0, it indicates the presence of power-law type tail with tail index $\alpha = p_0$. For example, if for $p = 1, 2, 3, 4$, the $R_n(p)$ does not converge to 0, it shows the absence of existence of any finite moment for the variable, which suggests that the variable has such a fat right tail that even the first theoretical moment is not finite. In such cases, we deal with such a heavy-tailed phenomenon that even the inferences from the sample average and standard deviation are meaningless. These plots can help us distinguish distributions belonging to class C\B from lognormal and class D\C. Distributions belonging to lognormal and class D\C have all their moments, while the distributions of class C\B may not, depending on the shape or tail parameter of the distribution. Hence, if the ratio R_n^p does not converge to 0 for any plotted p, no matter how many observations n are used, it proposes that no finite specific moments are likely to exist. For such cases, the distribution will belong to class C\B (regularly varying distributions).

5.4.6 Hill ratio plot

A generalization of the Hill estimator by Beirlant et al. (2006) is used in this study. This method is based on the fact that a_n is a consistent estimator of α if the tail is Pareto type (Class B) with tail index α. Let

$$a_n(x_n) = \frac{\sum_{i=1}^{n} I(X_i > x_n)}{\sum_{i=1}^{n} \log(X_i / x_n) I(X_i > x_n)}$$

(5.9)

Where, *I is an indicator function given as* $I(X_i > x_n) = \begin{cases} 1 & if \ X_i > x_n \\ 0 & if \ X_i < x_n \end{cases}$

In practice, one plots $a_n(x_n)$ as a function of x_n and looks for some stable region from which $a_n(x_n)$ can be considered as an estimator of α. In Equation (5.9) x_n is chosen to be large such that $P(X > x_n) \to 0$ and $nP(X > x_n) \to \infty$. The Hill estimator works very well in identifying the Pareto distributed data, but it becomes less effective for other regularly varying distribution functions (De Sousa and Michailidis, 2004; Cooke and Nieboer, 2011). For this study hill plot is used to identify the Pareto type tail (Class B); however it can also characterize the distribution of class D\C as the Hill ratio plot converges to zero for the subexponential tails that do not have power-law type tail (Ehsanzadeh et al., 2010).

5.4.7 Log-log plot

The log-log plot is the double logarithmic plot in which the log of the empirical survival function is plotted against the log of the ordered values of *u*. It is based on the fact that for exponential tail with mean θ empirical survival function is $\bar{F}(u) = P(X > u) = e^{-u/\theta}$ and for power-law tail with tail index α, empirical survival function is $\bar{F}(u) = P(X > u) = (u_0/u)^{\alpha}$. Taking the logarithm, we have for exponential type distribution

$$\log[P(X > u)] \approx -u/\theta$$

(5.10)

And for power-law distributions

$$\log[P(X > u)] \approx \alpha \log u_0 - \alpha \log u \approx C - \alpha \log u$$

(5.11)

Equation (5.11) shows a negative linear relationship between the logarithm of the survival function and the logarithm of *u*, suggesting that the tail probability for power-law (or regularly varying distributions, i.e., class C\B) can be identified in the presence of a linearly decreasing behavior of the plotted curve. In contrast, such behavior is not observed for other classes like the subexponential (class D\C) or exponential (class E). To improve interpretability of the plot, a naive linear fit of the decaying tail is shown with the red line in the log-log plots for different distributions in this study. Also, a test on the associated correlation coefficient is recommended to check the linearity of the curve. This plot helps discriminate the distributions belonging to class C from that of Class D\C (i.e., the distribution which belongs only to class D and not C). However, it cannot discriminate between subexponential and exponential classes.

5.5 *Proposed Comprehensive Decision Support System*

In the present study, a comprehensive DSS is proposed, which alleviates the limitations associated with the conventional DSS by incorporating advanced graphical methods in a well-defined order. A flow diagram depicting various methodological steps in the comprehensive DSS is provided in Figure 5.3.

Given a dataset, the first step of the proposed DSS is to construct the MEF, which can result in 3 cases, namely, the positive, null, and negatively sloped MEF plots representing the subexponential, exponential, and hyper-exponential tail types, respectively. To judge whether a dataset belongs to any of the three categories mentioned above, the confidence interval (CI) approach of MEF suggested by Nerantzaki and Papalexiou (2019) was considered. In this approach, we estimated the MEF slope from the entire dataset sample and tested for its significance against the null hypothesis of zero slope (indicating an exponential tail) by considering a 90% confidence interval. If the estimated slope lies within the CI, distribution is considered to be exponential, while based on the location of the estimated slope, either above the upper limit of CI or below the lower limit of CI, it can be categorized as subexponential or hyper-exponential, respectively. This algorithmically applied MEF approach eases the identification of the datasets and makes the process faster. Further, each case that might exist after the categorization is described in a detailed manner in the subsequent subsections.
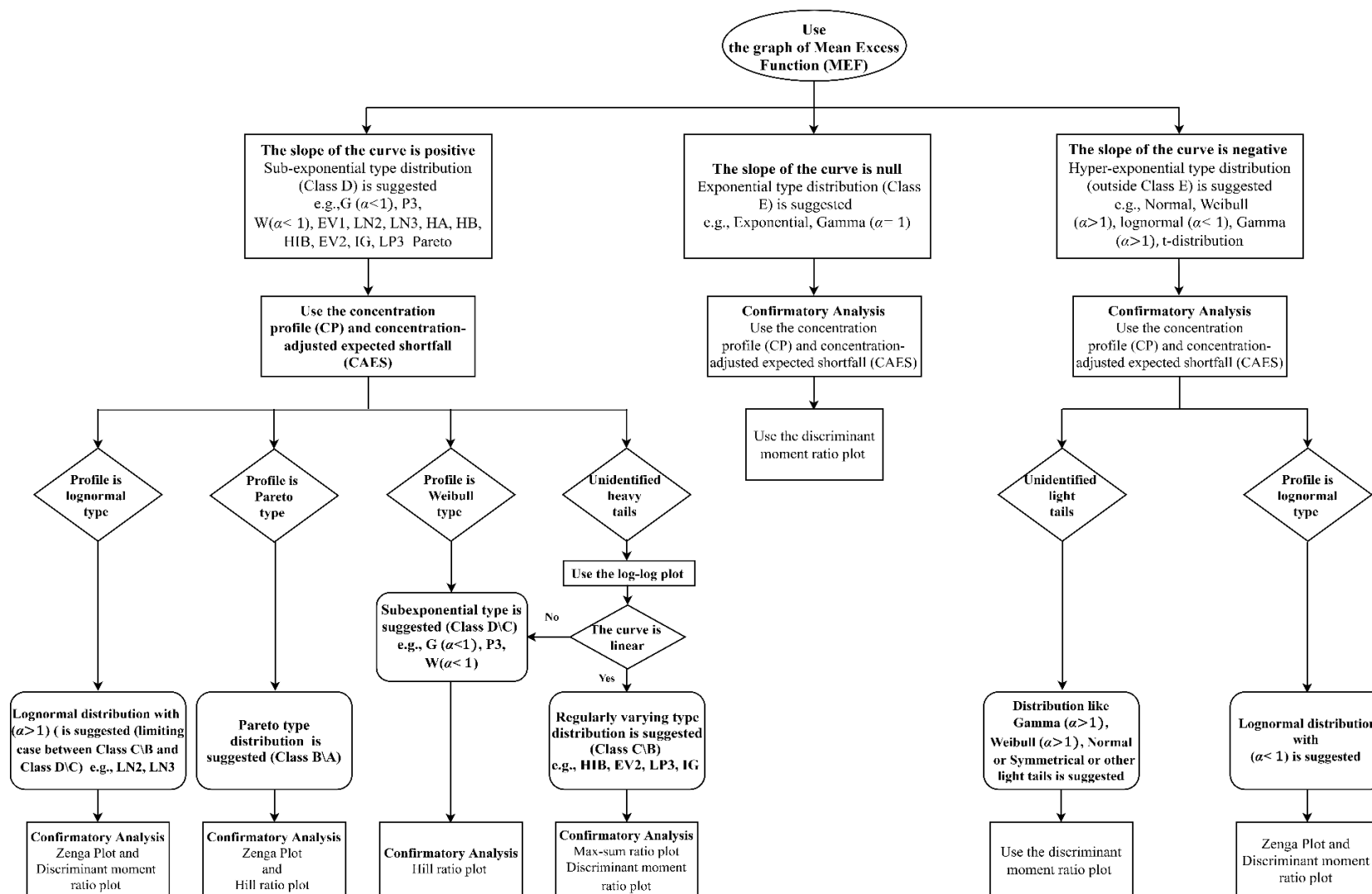
Figure 5.3 Steps involved in the proposed Decision Support System

5.5.1 Mean Excess Function: Positive slopes

This case indicates that the distribution of the dataset belongs to a Subexponential class, i.e., Class D. Further investigation is required to categorize the distribution in subclasses such as Pareto (class B\A), Lognormal distribution (LN), Regularly Varying (class C\B) and Subexponential (class D\C). The CP and CAES plots, when used in conjunction, help to subcategorize the Class D distributions. CP for each subclass of class D exhibits a non-increasing $G(\lambda)$ profile, with increasing truncation level (λ), in the range from 0 to 1. Each subclass can be identified based on the unique behavior of the CP plot, such as the slow decaying CP (i.e., LN and class C\B), faster decaying CP (i.e., class D\C), and constant CP (i.e., class B\A). Additionally, the CAES plot helps in confirming the identified subclass. Within a given subclass, the diverse behavior of CAES helps in confirming the subclass.

The distribution of the dataset is identified as a Lognormal distribution (a limiting case between class C\B and class D\C) based on the decreasing behavior of CP and the peculiar behavior of the CAES plot, both governed by the shape parameter of the distribution (α). For the LN class, CP has a starting point greater than 0.5 (i.e., $G(0) > 0.5$), when $\alpha > 1$ (representing the presence of heavy-tailed LN distribution) and the CAES plot tends to infinity as $\lambda \to 1$. Subsequently, confirmatory analysis for the LN distribution can be performed using the Zenga curve and DMR plot, as suggested by Cirillo (2013). For the Lognormal distribution, the Zenga curve illustrates a constant horizontal curve (i.e., $Z(u) = 1 - e^{-\alpha^2}$, $0 < u < 1$) corresponding to the variance of the LN distribution. The DMR plot can confirm the LN distribution based on the location of couple point $(\hat{\gamma}_2, \hat{\gamma}_3)$ within the lognormal zone.

The distribution of the dataset can belong to class B\A, i.e., Pareto type, when the CP plot detects a profile characterized by a continuous horizontal line whose intercept depends only on the shape parameter α. The Paretian CP is immune to any change with the increasing truncation level. Further, for any, the Parentian CAES always tends to infinity as $\lambda \to 1$ (i.e., $\lim_{\lambda \to 1} CAES_{\lambda} = \infty$). After the identification using CP and CAES, the confirmation of the Paretian tail can be accomplished based on the Zenga plot in addition to the conventionally preferred Hill ratio plot. An increasing Zenga curve and a stable Hill ratio plot around a non-null constant (i.e., an expected shape parameter) confirm that the datasets come from Class B\A. For the dataset to have a distribution belonging to class D\C, the CP has a faster rate of decay as compared to the LN class. Further, the CAES plot for these distributions tends to infinity, representing heavy-tailed distributions such as Weibull $\alpha < 1$. Followed by this, a non-stable Hill ratio plot can confirm the presence of class D\C.

It should be noted that the characterization of distributions into classes B\A and LN based on the simultaneous use of the CP and CAES plots can be achieved easily, as discussed above. However, the CP and CAES plots may not be able to discern between some distributions belonging to classes D\C and

C\B due to presence of mixed behavior. The CP for some distributions from class D\C (e.g., Gamma and Pearson type III distributions) exhibit behavior that can be represented by a mixture of CP for Weibull and exponential distributions. While the CP profile for distributions from class C\B (e.g., Inverted Gamma and Log-Pearson type III (LP3)) exhibits behavior that can be represented by a combination of CP for Pareto and Lognormal distributions. This situation arises due to the nested behavior of various classes of distributions. To resolve such a situation, firstly, the proposed DSS places these distributions under the category "unidentified heavy tails". Subsequently, a log-log plot is suggested to classify the dataset into either class D\C or class C\B. If the log-log plot is not linearly decreasing, then the distribution of the dataset from the unidentified heavy tails category can be assigned to the class D\C and confirmed through a Hill ratio plot which shows a convergence to 0 (confirmatory test for class D\C). On the contrary, if the slope of the log-log plot is linearly decreasing, then the distribution of the dataset can be associated with class C\B and confirmed through the MS plot and DMR plot (which are confirmatory tests for class C\B). In the MS plot, the maximum sum ratio ($R_n(p)$), is plotted against the number of observations n for various p statistical moments. A jump in $R_n(p)$ followed by non-convergence to zero above some threshold p_0 belonging to p indicates the presence of a power-law type tail with a tail index $\alpha = p_0$. It is to be noted that the $R_n(p)$ generally converges to zero for the distributions belonging to the lognormal and class D\C as they possess all moments. Contrary to this, the $R_n(p)$ does not converge to zero for the distributions from class C\B as higher moments do not exist for them. Due to this peculiar characteristic, the $R_n(p)$ is suggested to be used as a confirmatory test for the class C\B (i.e., regularly varying distributions). In the case of the DMR plot, the couple point depicting the C_v and C_s estimates of the Inverted Gamma distribution dataset lies on or close to the Inverted Gamma curve, while for other regularly varying distributions, such as the LP3 and Fréchet distributions, the point lies within the Gray zone representing the class C\B that are power tails or mixtures of lognormal and power tails and hybrid distributions.

5.5.2 *Mean Excess Function: Null slopes*

This kind of observation points towards class E, which comprises the Exponential distribution and Gamma distribution with $\alpha = 1$. Following this, a confirmatory analysis for class E can be carried out by using the CP and CAES plots. The CP for such a class starts around 0.5 and has a decreasing behavior (first convexly and then concavely) with a point of flex corresponding to truncation level $\lambda = 0.63$, while the CAES plots show a horizontal line (i.e., a constant slope). Further identification of Exponential distribution or Gamma distribution with $\alpha = 1$ can be accomplished based on the DMR plot, which has an exponential zone and gamma curve in it.

5.5.3 *Mean Excess Function: Negative slopes*

This kind of observation indicates the presence of a hyper-exponential class (Vela and Rodríguez, 2014) which encompasses the probability distributions such as Weibull ($\alpha > 1$), LN ($\alpha < 1$), Normal, Gamma ($\alpha > 1$), and other unidentifiable light tails in the proposed DSS. The CP and CAES plots can very well identify the profiles of light-tailed versions of Weibull or Lognormal distributions. In the case of light-tailed distributions, the CP plot starts below 0.5 ($G(0) < 0.5$); however, the CAES plot depicts different behavior for the candidate distribution. The CAES plot tends to zero for Weibull ($\alpha > 1$), Normal, Gamma ($\alpha > 1$), and Lognormal ($\alpha < 0.3$) distributions. In the case of lognormal distribution with $\alpha \in (0.3, 0.7)$, the CAES plot shows a unique *u*-shaped behaviour. Subsequently, the lognormal distribution can be verified through the Zenga and DMR plots. In the case of other distributions, the DMR plot can be used to categorize them into Gamma ($\alpha > 1$), Normal, or Weibull ($\alpha > 1$) distributions. The Normal and Gamma distributions have respective curves in the DMR plot, whereas the other hyper-exponential distributions, such as Weibull ($\alpha > 1$), can lie in the thin-tailed zone of the DMR plot.

5.6 *Efficacy evaluation of the proposed DSS through a simulation study*

In this section, the efficacy of the proposed DSS was evaluated through a simulation study. We generated the samples of different lengths from representative probability distributions belonging to various classes (e.g., D\C, C\B, B\A, etc.) and verified whether the proposed DSS could reclassify the samples into their respective classes or not. Herein, the samples of candidate distributions belonging to different classes, i.e., Pareto type I representing class B\A (i.e., Pareto type); Log Pearson type III and Inverted Gamma distributions representing class C\B (i.e., regularly varying distributions); 2-parameter lognormal distribution representing LN distribution with $\alpha > 1$ (i.e., LN limiting case between class C\B and D\C); Gamma distribution with $\alpha < 1$ and Weibull with $\alpha < 1$ representing class D\C (i.e., subexponential distributions); exponential distribution representing E\D (exponential type distributions); and Normal, Weibull with $\alpha > 1$, lognormal with $\alpha < 1$, and gamma with $\alpha > 1$ distributions representing the hyper-exponential type distributions (outside class E), are considered for the simulation experiment. The expressions for probability density functions (PDFs) of the aforementioned probability distributions are provided in Table 5.1 along with the sets of parameters considered for the data generation. The parameter values are selected based on past studies from the field of Hydroclimatology (Phien and Ajirajah, 1984; Nash, 1994; Reeve, 1996; Al-Zahrani and Husain, 1998; Aksoy, 2000; Heo et al., 2001; De Sousa and Michailidis, 2004; Sharma and Singh, 2010; Bhavana et al., 2012; Papalexiou et al., 2013; Cirillo, 2013; Martel et al., 2013; Sherif et al., 2014; Kozubowski et al., 2009; Mayooran and Laheetharan, 2014; Babu and Hooda, 2018; Farooq et al., 2018; Hussain et al., 2019; Ozonur et al., 2021; Gupta and Chavan, 2021; Gupta and Chavan, 2022). Sample sizes (*n*) of 500, 1000, 5000, 10000, 12000, 15000, 20000, and 25000 are considered for the simulation study. Once the samples (of a particular length) are generated from each of the distributions representing various classes (mentioned above), they are subjected to various steps in the proposed DSS. For each distribution, the plots of different graphical methods involved in DSS are presented in Figures 5.4 to

5.14. For brevity, the plots are presented for a typical sample size of 12000. Plots pertaining to other sample sizes are not shown here due to space constraint; however, they can be found in the supplementary material of Gupta and Chavan (2023b) from Figure S4 to S80. The subsequent subsections demonstrate the efficacy of the proposed DSS in characterizing the distributions from Sub-exponential type distribution (i.e., Class D), Exponential Type distributions (i.e., Class E), and Hyper-exponential type distributions (i.e., outside Class E).

Table 5.1 Expressions for Probability density function (*PDF*) along with the detail of parameter considered for data generation

Class of Distribution	Name of Probability Distribution	Probability density function	Parameter	Parameter Values
Hyperexponential (outside Class E)	Normal (*N*) (μ, σ)	$f_N = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[\dfrac{-1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$	mean (μ) standard deviation (σ)	(142.48, 60.726), (64.103, 26.717), (75.265, 41.898), (109.09, 61.071), (62.37, 23.45), (108.76, 49.56), (84.66, 28.99), (89.62, 1.72), (77.57, 27.73), (1789.04, 1332.44), (500, 1.2), (35.9534, 54.5892), (3.0599, 6.3873), (66.28, 17.81), (12665, 4710), (76.11, 27.73), (9.3, 21), (8.4, 22), (850.7, 84.13), (79.637, 47.942)
Hyperexponential (outside Class E)	Weibull (*W*) (α, β) When, $\alpha > 1$	$f_W(x) = \dfrac{\alpha}{\beta}\left(\dfrac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\dfrac{x}{\beta}\right)^{\alpha}\right)$	shape parameter $(\alpha > 0)$ scale parameter $(\beta > 0)$	(1.491, 52.72), (17.7532, 31.5581), (19.2964, 30.8359), (2.9114, 13900), (12.07339, 3.059887), (1.3138, 99.3311), (2.2928, 164.4540), (1.9958, 1149.90), (3.3485, 154.8), (2.6347, 71.503), (2.3011, 82.972), (2.8985, 110.24), (1.4832, 87.627), (2.35, 5.63), (2.69, 6.41), (3.15, 6.65), (2.61, 6.47), (2.44, 9.70), (3, 2), (12.16, 887.4)
Hyperexponential (outside Class E)	Lognormal (*LN2*) (α, β) When, $\alpha < 1$	$f_{LN}(x) = \dfrac{1}{\sqrt{\pi}\alpha x} \exp\left(-\ln^2\left(\dfrac{x}{\beta}\right)^{\frac{1}{\alpha}}\right)$	shape parameter $(\alpha > 0)$ scale parameter $(\beta > 0)$	(0.088294, 1.0737), (0.1, 0.1), (0.1878, 1.5062), (0.2,0.25), (0.3,0.5), (0.4562, 10.7711), (0.025,4.234), (0.35, 9.6792), (0.31, 1000.22), (0.16, 100), (0.29, 19.162), (0.39, 100.5), (0.13, 40.314), (0.43, 5), (0.3704, 4.8866), (0.4617, 4.0644), (0.272, 4.1881), (0.30778, 4.7685), (0.46, 5.158), (0.4, 121.58)

Table 5.1 (continued…..)

Class of Distribution	Name of Probability Distribution	Probability density function	Parameter	Parameter Values
Hyperexponential (outside Class E)	Gamma (G) (α, β) When, $\alpha > 1$	$f_G(x) = \dfrac{1}{\beta\,\Gamma(\alpha)}\left(\dfrac{x}{\beta}\right)^{\alpha-1} \exp\left(-\dfrac{x}{\beta}\right)$	shape parameter ($\alpha > 0$) scale parameter ($\beta > 0$)	(1.737, 17.3272), (2.433, 120), (7.0028, 30.1988), (2.1743, 18.1197), (10.6443, 32.7686), (8.743, 97.30), (1.8, 28.18), (2.433, 120), (5.5048, 25.882), (5.7569, 11.135), (3.227, 23.323), (3.191, 34.19), (4.8157, 22.584,), (2.3215, 42.916), (6.572, 50.637), (8.195, 38.181), (3.787, 396.11), (2.122, 33.669), (1.317, 46.715), (2.5411, 20.484)
Class E	Exponential (E)	$f_E(x) = \alpha \exp(-x\alpha)$	shape parameter ($\alpha > 0$)	(0.001), (0.007), (0.0012), (0.0156), (0.0133), (0.017), (0.09), (0.1), (0.11), (0.12), (0.5), (0.75), (0.8), (1), (1.5), (2), (3), (5), (7.5) (10)
Class D\C	Gamma (G) (α, β) When, $\alpha < 1$	$f_G(x) = \dfrac{1}{\beta\,\Gamma(\alpha)}\left(\dfrac{x}{\beta}\right)^{\alpha-1} \exp\left(-\dfrac{x}{\beta}\right)$	shape parameter ($\alpha > 0$) scale parameter ($\beta > 0$)	(0.0379, 3.4304), (0.092, 17.50), (0.010, 3.79,), (0.219, 23.15), (0.294, 28.18), (0.9060, 17.6201), (0.1139, 5.1253), (0.731, 18.3819), (0.5, 17.50), (0.7039, 23.7079), (0.5913, 22.0360), (0.5934, 17.7122), (0.6373, 121.9516), (0.681, 16.0844), (0.7268, 24.5152), (0.6957, 18.7077), (0.0443, 11.27), (0.561, 10.23), (0.436, 11.05), (0.495, 6.892), (0.341, 16.87), (0.375, 16.72), (0.569, 13.65), (0.023, 14.760), (0.075, 31.765), (0.084, 12.16), (0.134, 1.907), (0.10, 1.053), (0.231, 5.749), (0.608, 23.594)

Table 5.1 (continued…..)

Class of Distribution	Name of Probability Distribution	Probability density function	Parameter	Parameter Values
Class D\C	Weibull (*W*) (α,β) When, $\alpha<1$	$f_W(x)=\dfrac{\alpha}{\beta}\left(\dfrac{x}{\beta}\right)^{\alpha-1}\exp\left(-\left(\dfrac{x}{\beta}\right)^{\alpha}\right)$	shape parameter $(\alpha>0)$ scale parameter $(\beta>0)$	(0.3, 20), (0.23, 0.02), (0.661, 4.33), (0.678, 5.91), (0.692, 6.88), (0.3087, 0.3476), (0.90, 12.2075), (0.8501, 12.4927), (0.3066, 0.3850), (0.7859, 0.1224), (0.2012, 0.6952), (0.8655, 12.2138), (0.8158, 12.6215), (0.8715, 6189.6259), (0.6952, 0.2012), (0.6976, 0.2321), (0.0469, 0.7360), (0.434, 7.71), (0.8044, 0.1088), (0.7732, 0.1556), (0.21209, 1066.1), (0.23585, 851.73), (0.722, 2.199), (0.3, 4), (0.68, 16.599), (0.7, 12.778), (0.82, 12.042), (0.13, 111.41), (0.797, 42.691), (0.8554, 18.3)
LN (a limiting case between Class C and D\C)	Lognormal (*LN2*) (α,β) When, $\alpha>1$	$f_{LN}(x)=\dfrac{1}{\sqrt{\pi}\,\alpha x}\exp\left(-\ln^2\left(\dfrac{x}{\beta}\right)^{\frac{1}{\alpha}}\right)$	shape parameter $(\alpha>0)$ scale parameter $(\beta>0)$	(1.6097, 1.52), (1.060, 8.78), (1.087, 9.46), (1.107, 10.59), (1.1873, 10.0345), (1.2208, 10.2748), (1.2814, 9.1881), (1, 10), (1.3141, 9.6792), (1.5, 15), (1.8924, 2.6398), (2, 20), (2.115, 24.5317), (2.2519, 24.0319), (2.284, 76.74), (3, 25), (3.5, 3.0599), (1.5023, 1.9568), (1.4449, 1.6966), (1.4704, 2.0529), (1.1634, 2.1468) , (1.6223, 3.3920), (1.1618, 1.3093), (2.6582, 4.3432), (2.44, 4.3794,), (1.186, 2.137,), (1.198, 2.001), (1.118, 2.123), (1.3192, 28.7343), (3.6, 32.546)

Table 5.1 (continued…..)

Class of Distribution	Name of Probability Distribution	Probability density function	Parameter	Parameter Values
Class C\B	Inverse Gamma (*IG*) (α, β)	$f_{IG}(x) = \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x} \mathrm{I}(x>0)$ *where,* $\Gamma(\alpha)$ *is the gamma function* $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$ $X \sim Ga(\alpha, \lambda),$ *where* λ *is the rate parameter* $Y = \dfrac{1}{X} \sim IG(\alpha, \lambda)$	shape parameter ($\alpha > 0$) scale parameter ($\beta > 0$)	(1, 0.1), (3, 1), (12, 2), (160, 12), (0.8, 17.50), (0.9, 50), (1, 23.15), (1.1, 1), (1.5, 1), (1.9, 1), (2, 15), (4, 94.72), (5, 15), (6, 17), (7, 3), (1.6, 0.2), (10, 0.6), (30, 1.2), (1.2, 0.3), (2, 1)
Class C\B	Log Pearson type III (*LP3*) (α, β, γ)	$f_{LP3} = \dfrac{1}{x \,|\beta|\, \Gamma(\alpha)} \left(\dfrac{\ln(x)-\gamma}{\beta} \right)^{\alpha-1} \exp\left(-\dfrac{\ln(x)-\gamma}{\beta} \right)$	shape parameter ($\alpha > 0$) scale parameter ($\beta \neq 0$) lower bound or location parameter (γ)	(41.62, 0.12, 5.17), (47.56, 0.15, 3.86), (6.41, 0.28, 3.67), (36.73, 0.15, 0.48), (41.62, 0.15, 3.18), (330.58, 0.04, 8.23), (36.73, 0.08, 0.34), (3.56, 0.41, 4.41), (16, 0.125, 6), (82.6, 0.209, 2.82), (90.70, 0.16, 3.33), (11.053, 0.21461, 4.5936), (1.2845, 0.41788, 5.5422), (8.91, 0.130, 3.73), (15.219, 0.1357, 6.2541), (6.6347, 0.32385, 5.6431), (1111.11, 0.07, 2.52), (400, 0.05, 2.33), (34.6, 0.17, -1.46), (4444.44, 0.02, 6.27)

Table 5.1 (continued…..)

Class of Distribution	Name of Probability Distribution	Probability density function	Parameter	Parameter Values
Class B	Pareto I (P) (β, α)	$f_P(x) = \dfrac{\alpha\beta^\alpha}{x^{\alpha+1}}$	Shape or tail parameter ($\alpha > 0$) scale parameter ($\beta > 0$)	(51.5, 1.0582), (10, 2), (1, 1), (1, 1.4), (1, 1.9), (1, 2.4), (10, 2.5), (100, 1.5), (34.241, 2.38), (0.7920, 3.03), (7.54, 0.134), (8.8, 0.14), (9.51, 0.145), (10, 5), (7.6815, 4.27), (8.1352, 3.97), (7.5730, 3.759), (8.1501, 3.51), (25.6076, 1.44), (31.89, 1.34), (54.79, 0.798), (100, 2.6), (1, 2.003), (51.5, 1.02), (12.1, 1.57), (7.1, 2.22), (1.5, 1.17), (1, 1.42), (8.1352, 0.5), (25.6076, 0.6910)
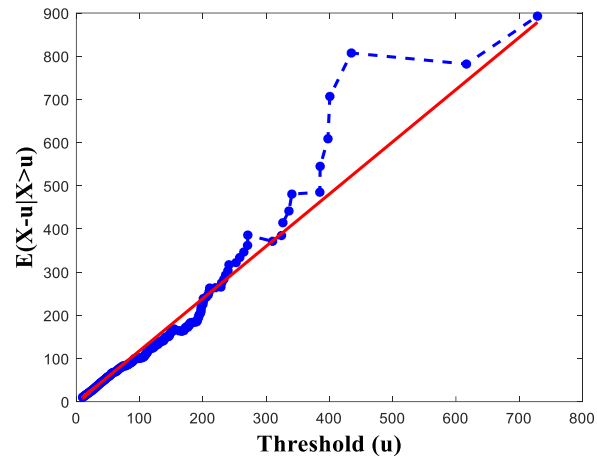
(i) MEF
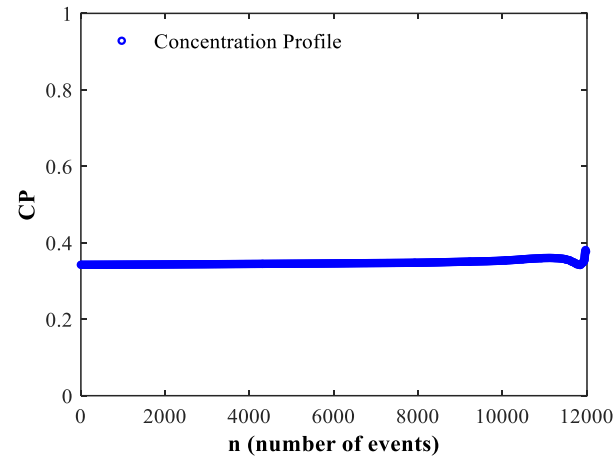
(ii) CP
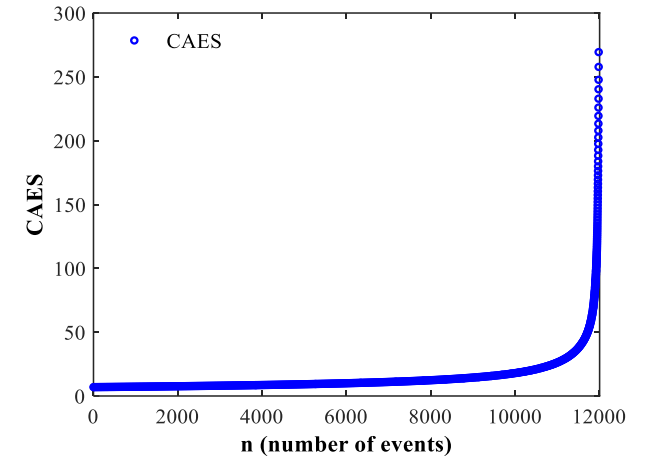
(iii) CAES
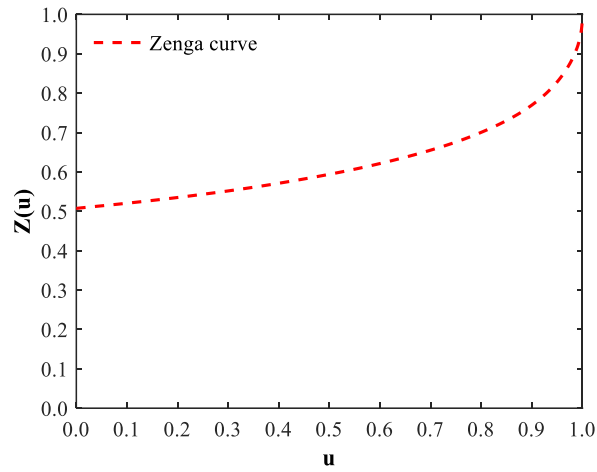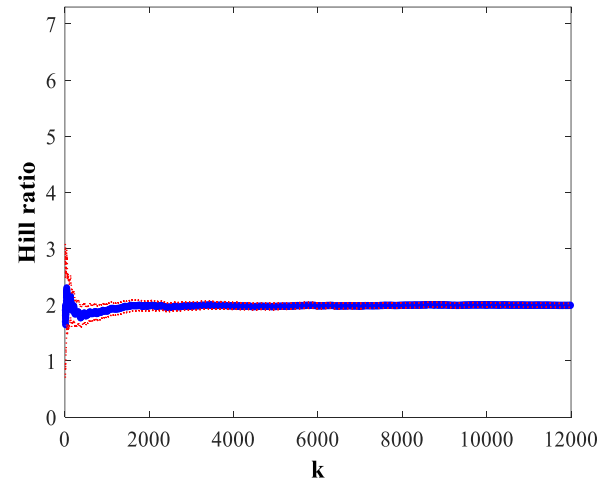
(iv) Zenga plot

(v) Hill ratio plot

Figure 5.4 MEF, CP, CAES, Zenga, and Hill ratio plot for Pareto I distribution $(\beta = 10, \alpha = 2)$ with sample size $n = 12000$

106
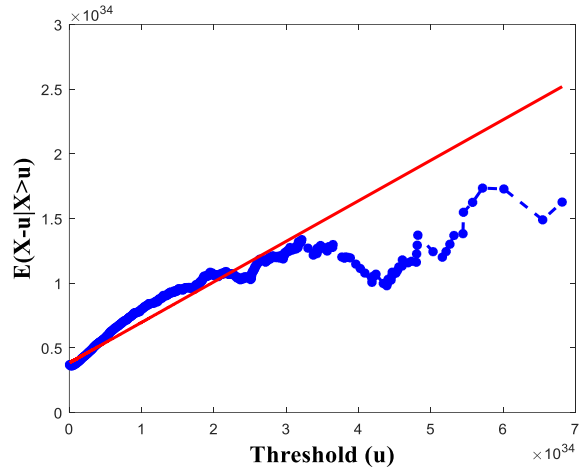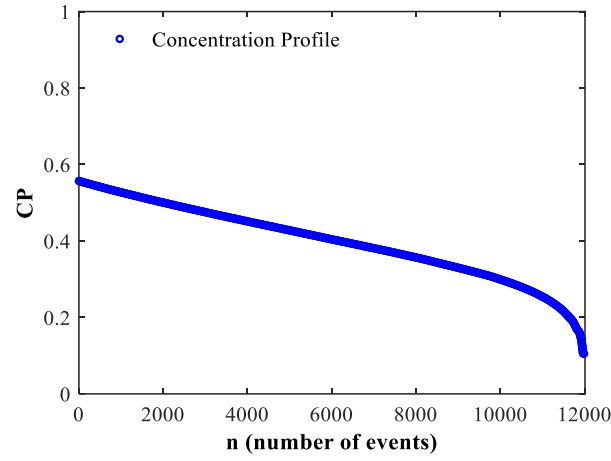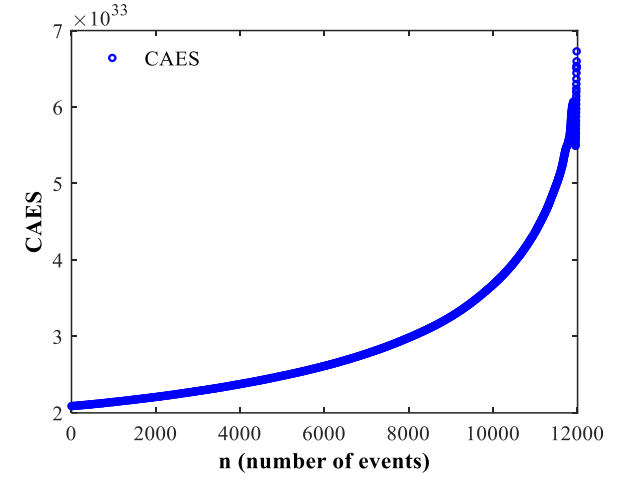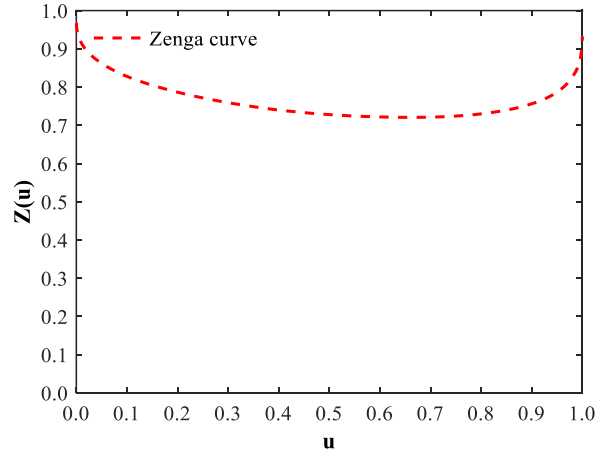
(i) MEF

(ii) CP

(iii) CAES

(iv) Zenga plot
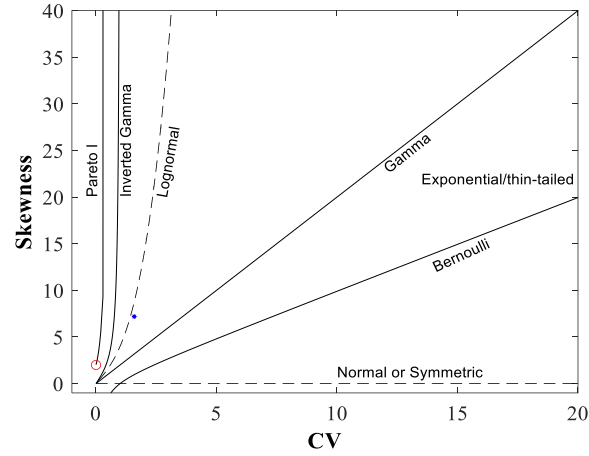
(v) DMR plot

Figure 5.5 MEF, CP, CAES, Zenga, and DMR plot for Lognormal distribution $(\alpha = 1.107, \beta = 10.59)$ with sample size $n = 12000$
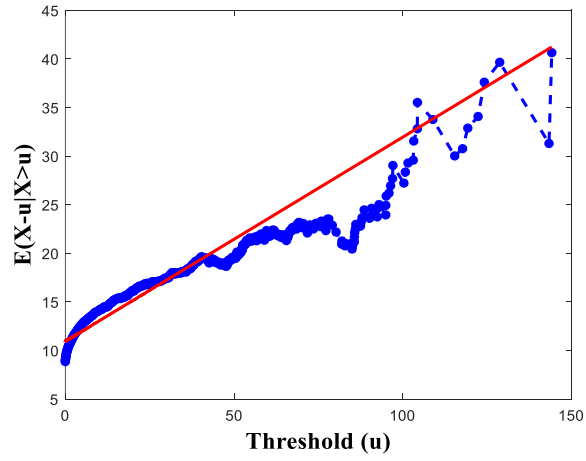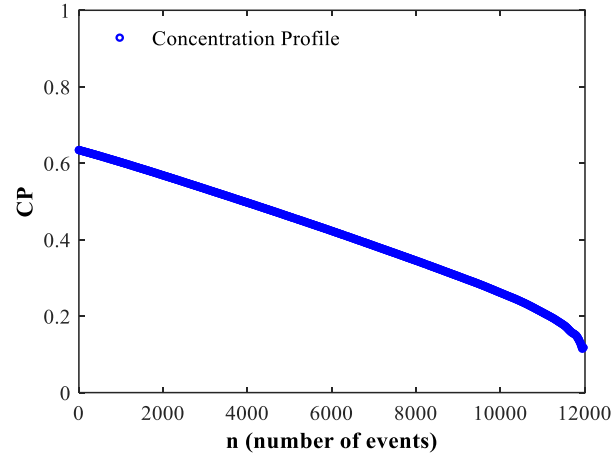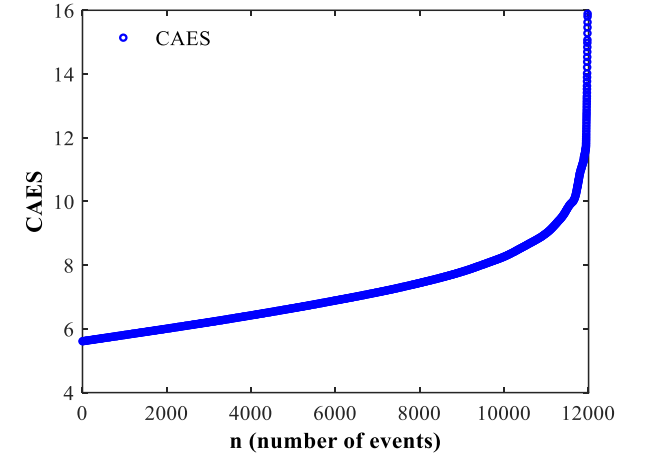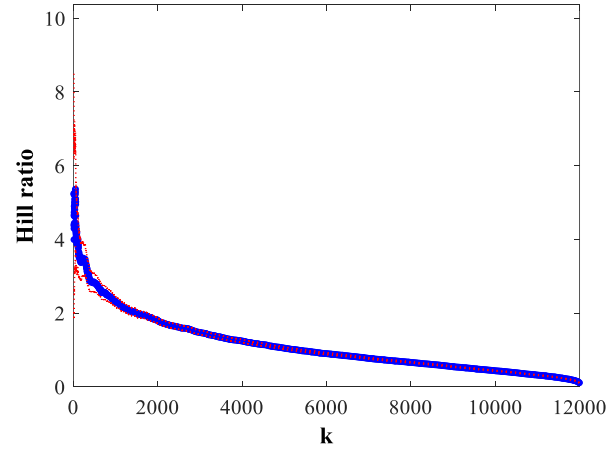
(i) MEF

(ii) CP

(iii) CAES

(iv) Hill ratio plot

Figure 5.6 MEF, CP, CAES, and Hill ratio plot for Weibull distribution $(\alpha = 0.692, \beta = 6.88)$ with sample size $n = 12000$
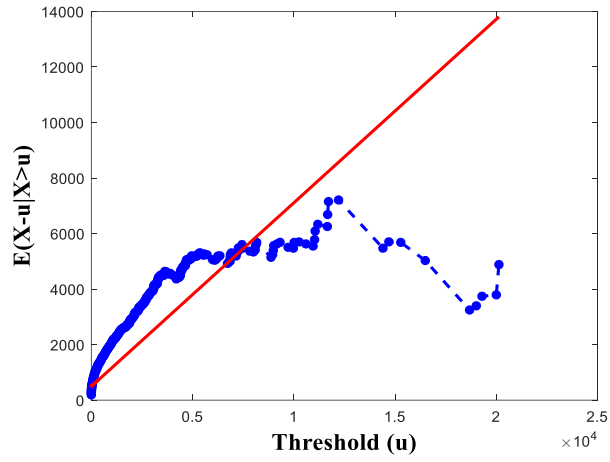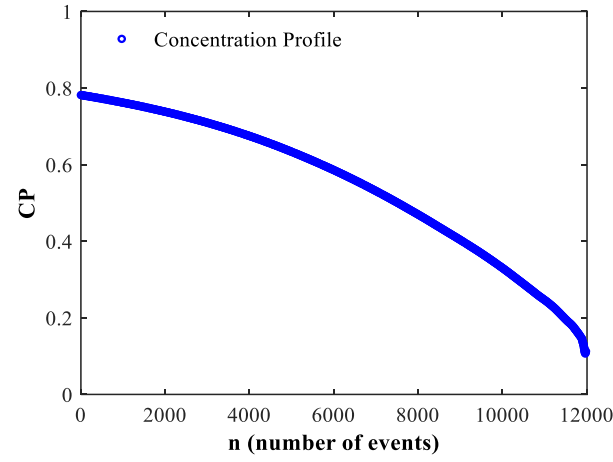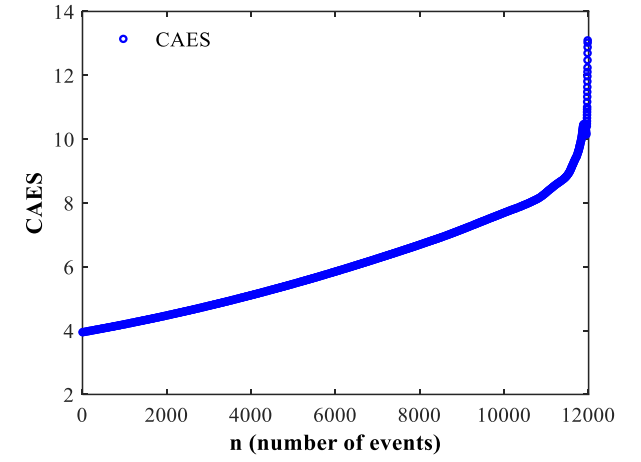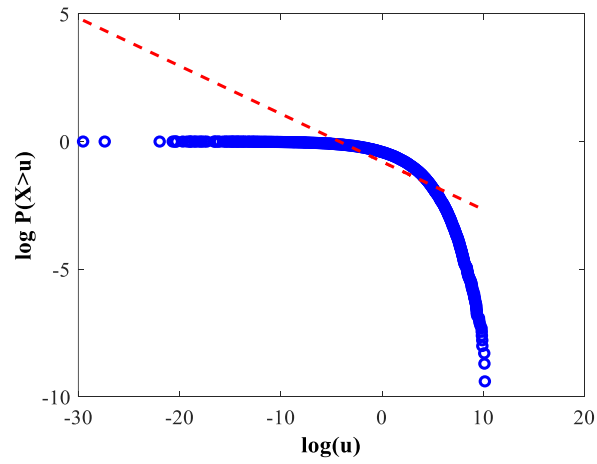
(i) MEF **(ii) CP** **(iii) CAES**
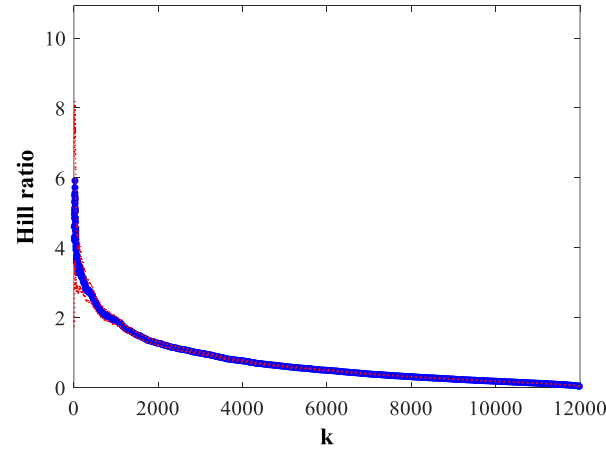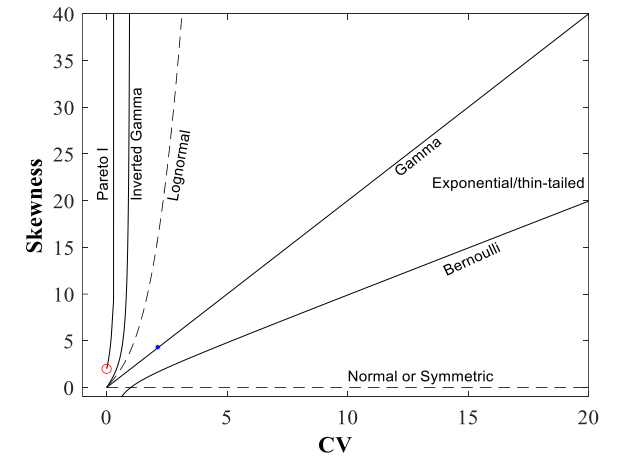
(iv) Log-log plot **(v) Hill ratio plot** **(vi) DMR plot**

Figure 5.7 MEF, CP, CAES, log-log, Hill ratio, and DMR plot for Gamma distribution $(\alpha = 0.219,\ \beta = 23.15)$ with sample size $n = 12000$
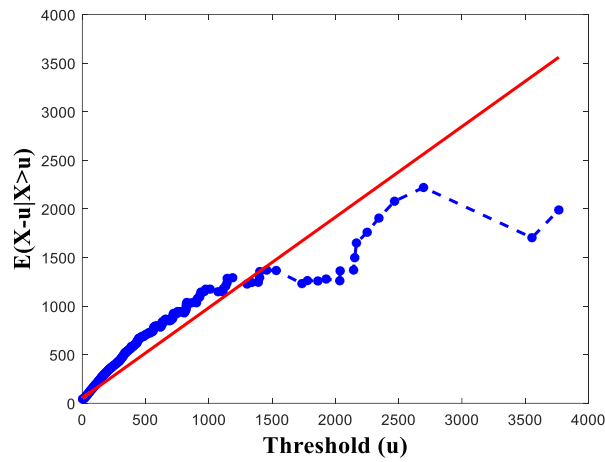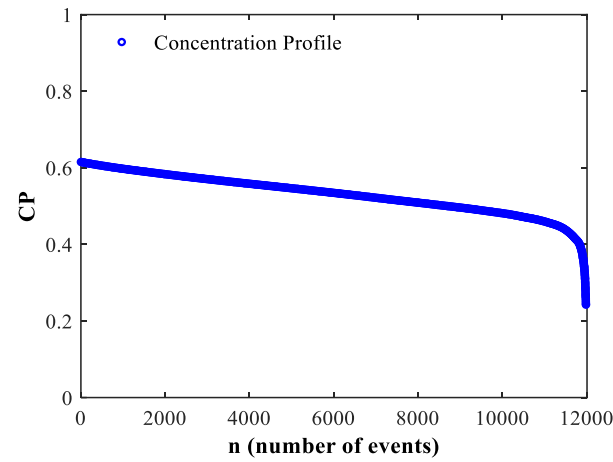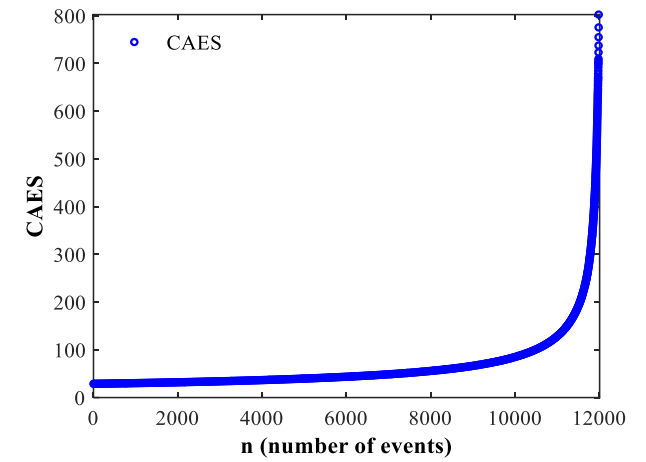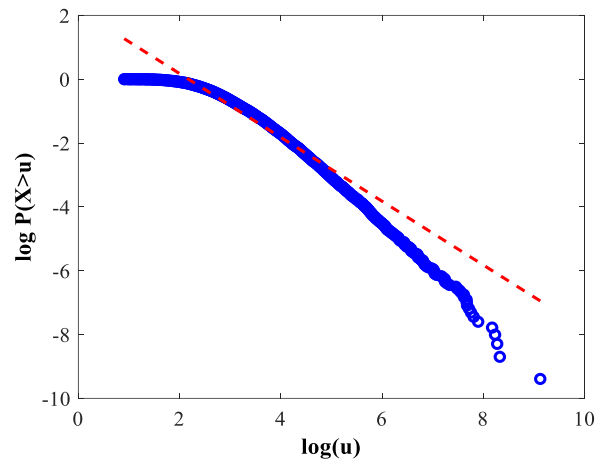
(i) MEF

(ii) CP

(iii) CAES
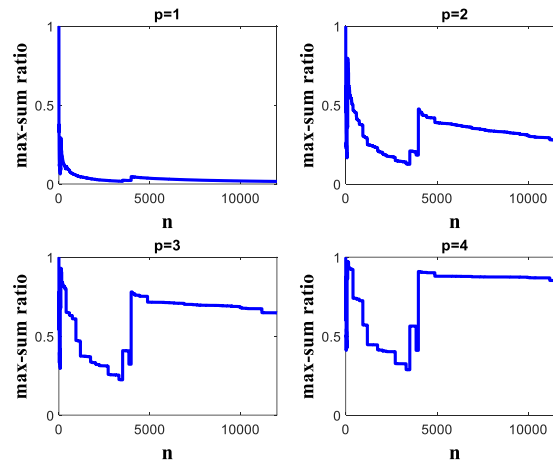
(iv) Log-log plot
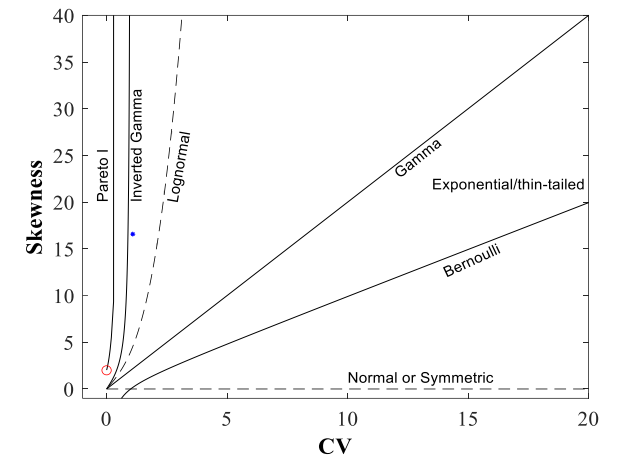
(v) MS plot

(vi) DMR plot

Figure 5.8 MEF, CP, CAES, log-log, MS, and DMR plot for Inverted Gamma distribution ($\alpha = 2$, $\beta = 15$) with sample size $n = 12000$
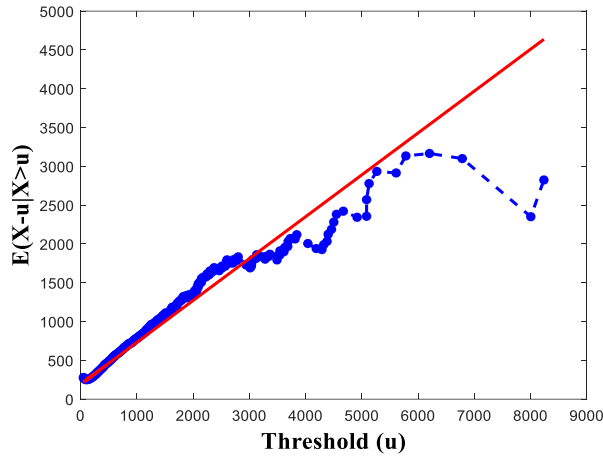
(i) MEF
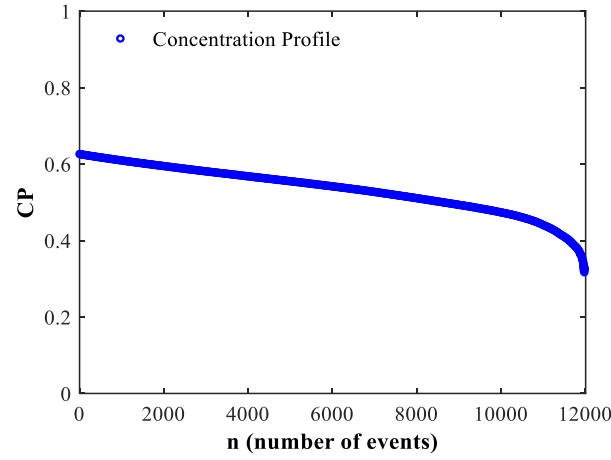
(ii) CP
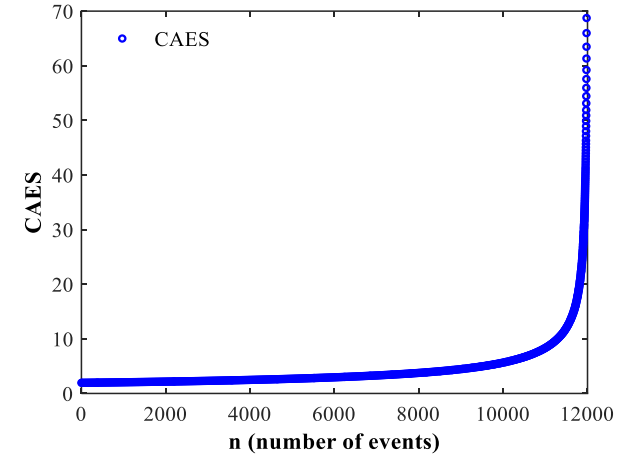
(iii) CAES
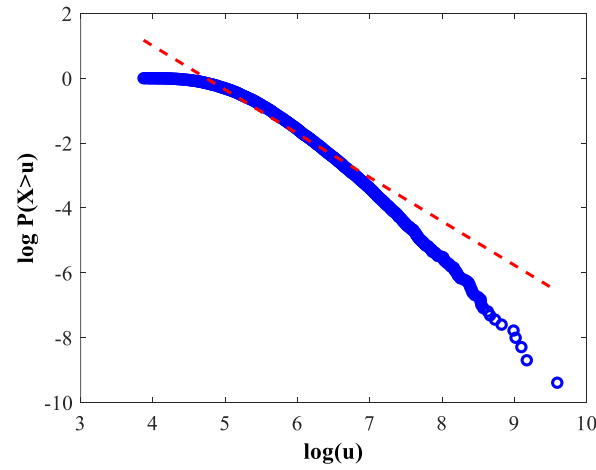
(iv) Log-log plot
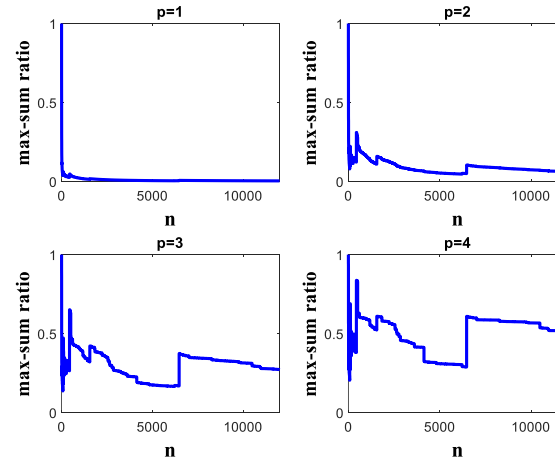
(v) MS plot
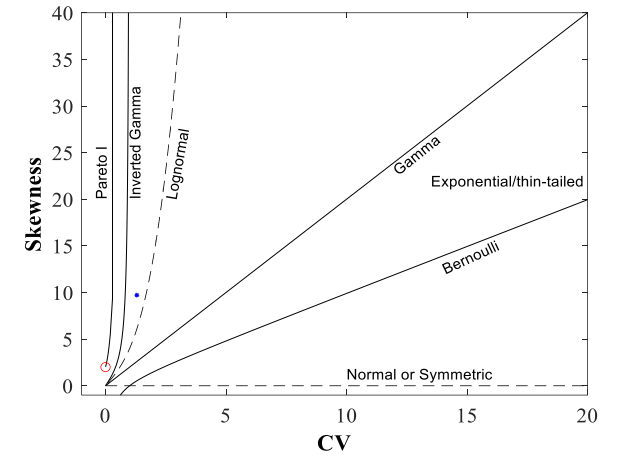
(vi) DMR plot

Figure 5.9 MEF, CP, CAES, log-log, MS, and DMR plot for Log-Pearson type III distribution $(\alpha = 6.41, \ \beta = 0.28, \ \gamma = 3.67)$ with sample size n=12000

5.6.1   Subexponential distribution type

The proposed DSS was applied to the generated datasets from various distributions. Firstly, the MEF plot, which is a popular choice for identifying Subexponential distributions, is applied to a dataset. The slope of MEF plots for distribution like Pareto, Lognormal, Weibull, Gamma, Inverted Gamma, and Log-Pearson type III distribution is found to be linear and positive (see Figures 5.4 (i), 5.5 (i), 5.6 (i), 5.7 (i), 5.8 (i), 5.9 (i)). The outcome of the algorithmic approach of MEF indicated the Sub-exponential tails for these samples as the estimated MEF slopes of these samples lie above the upper limit of the 90% CI of the Exponential tail. Identification of the subexponential class as a whole is done with ease using the algorithmic approach of the MEF plot. However, the distinction of Pareto (Class B\A), regularly varying (class C\B), LN (limiting case between Class C\B and D\C), and only subexponential (class D\C) required more sophisticated tools for classification.

Firstly, the proposed DSS was applied to the datasets generated from 30 sets of Pareto type distribution (i.e., class B\A). The CP showed continuous horizontal lines for all 30 sets considered for simulation. In addition, it was observed that the $G(0)$ increases with the decrease in the shape parameter (α) of the Pareto distributions. The smaller values of α (i.e., higher constant level of $G(\lambda)$) correspond to heavier and riskier Paretian tails. Further, the CAES plots for all sets of Pareto distribution indicate an increasing curve as $\lambda \to 1$. This observation is in line with the previous studies (Fontanari et al., 2018a). A constant CP and infinitely increasing CAES for Pareto distribution indicate the sustenance of peaks in the Parentian tails. Finally, an increasing Zenga curve and a stable Hill ratio plot at a non-null constant for the generated data confirm that the data comes from Class B\A (see Figures 5.4 (iv) and 5.4 (v)). Secondly, the application of the proposed DSS to datasets generated from 30 sets of LN distribution with $\alpha > 1$ indicated decreasing CP with a flatter profile having and a diverging CAES (i.e., $\lim_{\lambda \to 1} CAES_\lambda = \infty$) (refer to Figures 5.5 (b) and 5.5 (c)). This suggested that the datasets belong to the LN distribution class with $\alpha > 1$. The confirmation of the same was achieved based on the constant Zenga curve (except at the extremities), and the couple point $(\widehat{\gamma}_2, \widehat{\gamma}_3)$ plotted in the lognormal zone close to the lognormal curve in the DRM plot. The CP, CAES, and Zenga plots not only help in identifying the lognormal distributions correctly but also give some idea about their shape parameters. A CP with $G(0) > 0.5$ lying uniquely above the exponential CP, a diverging CAES, and a Zenga curve with a high value (i.e., $Z(u) \approx 0.767$) altogether suggested a large value, preferably above 1 for the shape parameter α, which is true in this case. Thirdly, the DSS was applied to 30 datasets generated from Weibull distribution with shape parameter $\alpha < 1$, which is a candidate of class D\C. Figures 5.6 (i), (ii), and (iii) show the characterization of one typical dataset with the help of MEF, CP, and CAES plots. The figures illustrated a decreasing CP lying uniquely above exponential CP and diverging CAES tending to infinity as the truncation level increases. Figure 5.6 (iv) shows that the curve of the hill ratio plot tends to zero, which confirms that the underlying distribution of the generated data belongs to class

D\C. Similar observations were noted for the remaining 29 sets of parameters belonging to the Weibull distribution with $\alpha < 1$.

Finally, it should be noted that some distributions belonging to classes D\C and C\B might not be discernible directly based on CP and CAES methods as they are represented by a combination of distributions. Such distributions are categorized as "unidentified heavy tails" in the proposed DSS. Figures 5.7 (ii), 5.8 (ii), and 5.9 (ii) show the CP plots for Gamma distribution from class D\C and Inverted Gamma (IG) as well as Log-Pearson type III (LP3) from class C\B, respectively. It should be noted that the CP for the candidate distributions from class C\B tends to be flatter than the lognormal distribution, while it is less flat relative to the Pareto distribution (class B\A), for which the CP is a horizontal line. The flatter the CP, the higher is the risk of recurrence of the extreme event within the tails of the distribution. To categorize the unidentified heavy tails, the DSS proposes to use the log-log plot to discern the distributions of classes C\B and D\C. Figure 5.7 shows the characterization of the dataset generated from the Gamma distribution with shape parameter $\alpha < 1$, which is a candidate of class D\C. A non-linear decreasing slope of the log-log plot was observed, which is a characteristic of D\C class, and the same was subsequently confirmed through the Hill ratio plot(showing convergence to 0). Similar observations are noted for all 30 sets of parameters belonging to the Gamma distribution with representing the class D\C. Figures 5.8 and 5.9 present the characterization of datasets generated from IG and LP3 distributions from class C\B, respectively. The log-log plot showed a linear profile for both datasets indicating that they belong to class C\B. To confirm the class C\B, MS plot, and DMR plot were constructed to categorize the datasets as shown in Figures 5.8 (v) and 5.9 (v). It can be noted from the figures that the MS plot does not converge to zero for $p \geq 1$ ($p \geq 2$) in the case of IG(LP3) distribution which confirms that the datasets belong to class C\B. Further, the DMR plot confirms the class C\B through plotting C_v and C_s estimates (of the datasets) on or close to the IG curve. Figure 5.8 (f) shows that the point for the generated data from IG falls on or close to the IG curve, while Figure 5.9 (f) illustrates that the generated data from LP3 lies within the Gray zone.

5.6.2 *Exponential type distribution*

The proposed DSS was applied to the datasets generated from Exponential type distributions. Figure 5.10 presents the DSS-based characterization of a typical dataset. The MEF plot for the dataset visually showed a null or constant slope indicating the presence of Exponential type distribution (i.e., class E). This was also found based on the algorithmic approach of MEF, where the estimated MEF slope lies within the 90% confidence interval (CI) of MEF slopes resulting from the Exponential tail. This was further confirmed by the CP and CAES plots. The CP plot showed that $G(0)$ starts around 0.5, with a decreasing profile (first convexly and then concavely) having a flex point around 7500 sample length value (corresponding to truncation level $\lambda = 0.625$). The CAES plots showed a horizontal line (i.e., a constant slope), which confirmed that the dataset belongs to class E. Further identification of

Exponential distribution was accomplished based on the location of the couple point $(\hat{\gamma}_2, \hat{\gamma}_3)$ in the Exponential zone of the DMR plot. Similar observations are noted for the remaining 19 sets of parameters belonging to the Exponential distribution representing class E.

5.6.3 *Hyper-exponential type distribution*

Finally, the effectiveness of the proposed DSS is assessed for characterization of the hyper-exponential type distributions (outside class E) (e.g., normal, weibull ($\alpha > 1$), gamma ($\alpha > 1$), and lognormal ($\alpha < 1$)). Figures 5.11 (i), 5.12 (i), 5.13 (i), and 5.14 (i) showed negative slopes for the MEF plots of all these distributions. The algorithmic approach of MEF revealed that the generated datasets belong to the hyper-exponential type distributions as the estimated MEF slopes of these samples lie below the lower limit of CI for the Exponential tail. Further, the hyper-exponential type distributions were confirmed using the CP and CAES plots (see Figures 5.11 (ii), 5.11 (iii), 5.12 (ii), 5.12 (iii), 5.13 (ii), 5.13 (iii), and 5.14 (ii)). The CP plots for the datasets indicated a fast-decreasing behavior with $G(0) < 0.5$. Also, the CAES plots showed rapid decay to zero for datasets generated from Weibull ($\alpha > 1$), lognormal ($\alpha < 0.3$), and other light-tailed distributions. CAES profile for the dataset generated from a lognormal distribution with $\alpha \in (0.3, 0.7)$ was found to have a unique *u*-shape behavior (refer to Figure 5.14 (iii)). Further identification of individual distributions (i.e., Normal, Weibull, and Gamma) from this class was performed by using the DMR plot. The DMR plots of all these distributions are shown in Figures 5.11 (iv), 5.12 (iv), and 5.13 (v). In the case of lognormal ($\alpha < 1$) distribution, the Zenga plot and the DMR plot were considered for identification (see Figures 5.14 (iv) and 5.14 (v)).
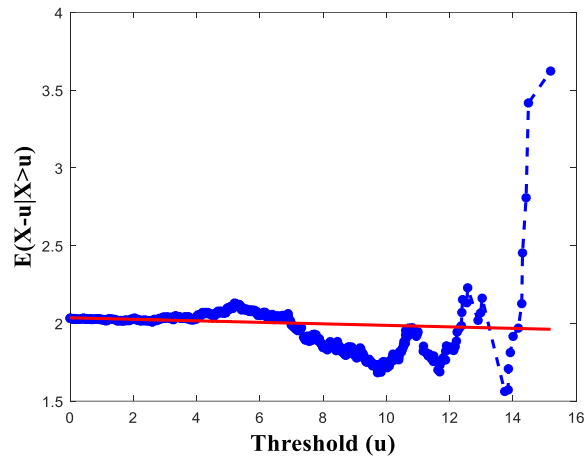
(i) MEF
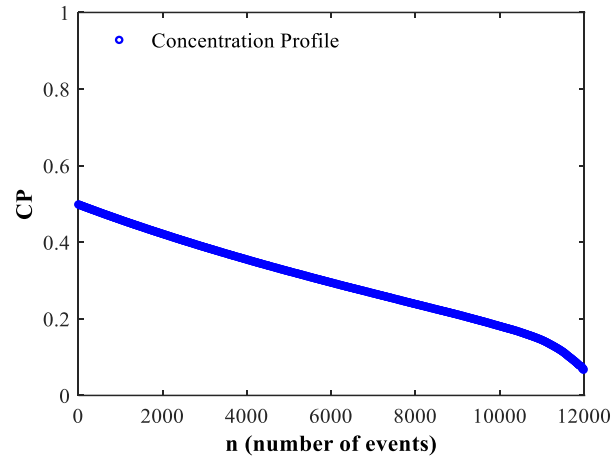
(ii) CP
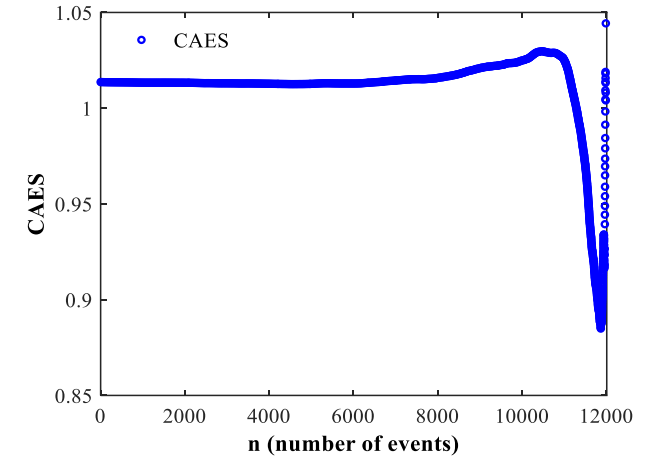
(iii) CAES
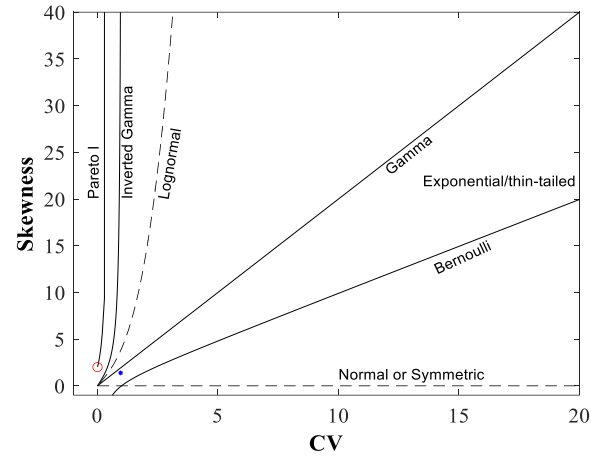
(iv) DMR

Figure 5.10 MEF, CP, CAES, and DMR plot for Exponential distribution ($\alpha = 0.5$) with sample size $n = 12000$
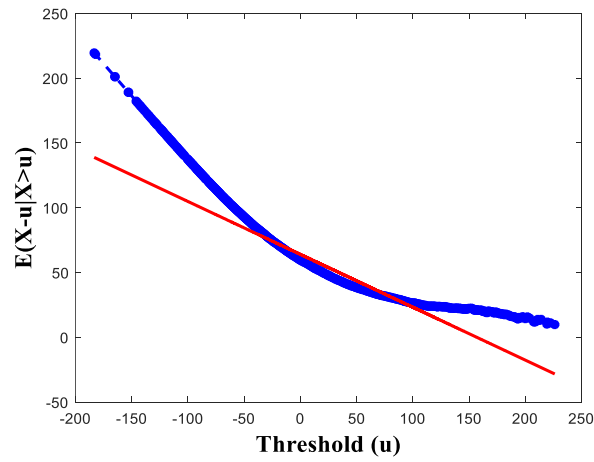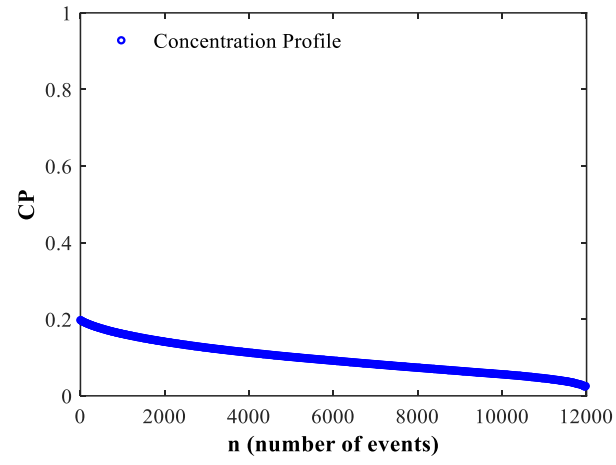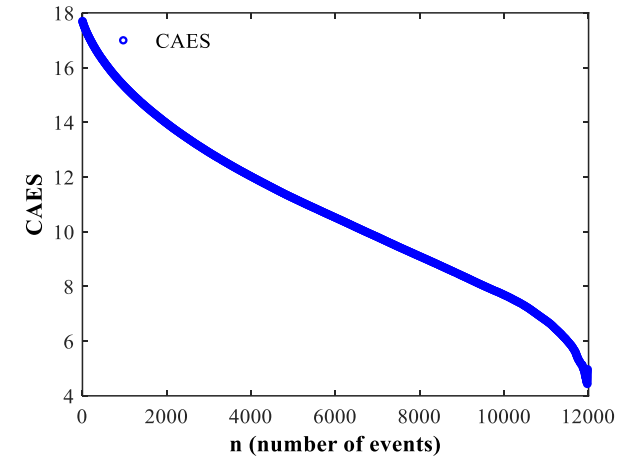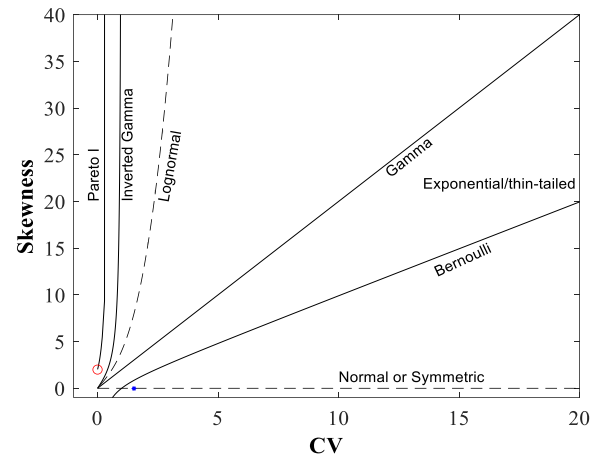
(i) MEF

(ii) CP

(iii) CAES

(iv) DMR

Figure 5.11 MEF, CP, CAES, and DMR plot for Normal distribution $(\mu = 89.62, \sigma = 31.72)$ with sample size $n = 12000$
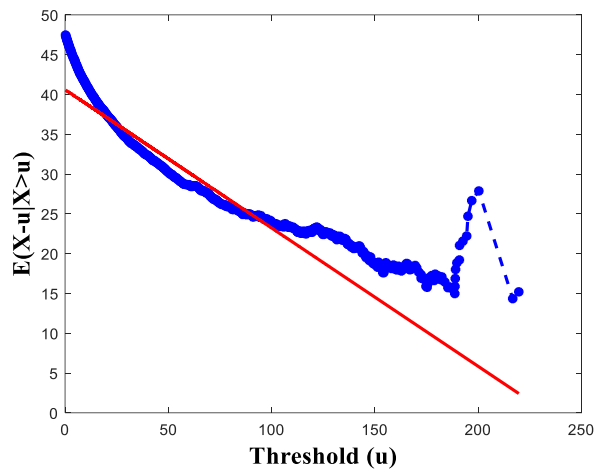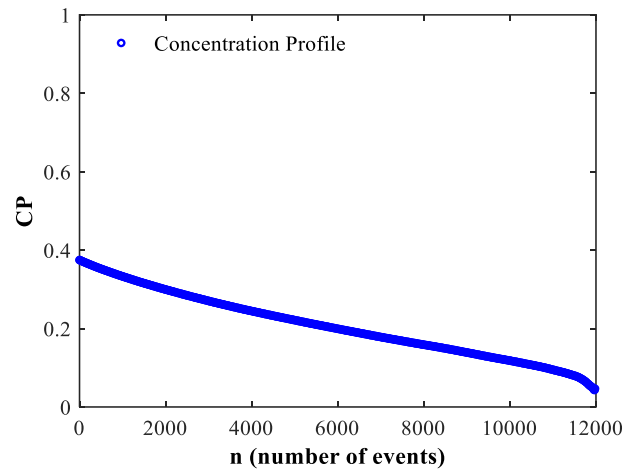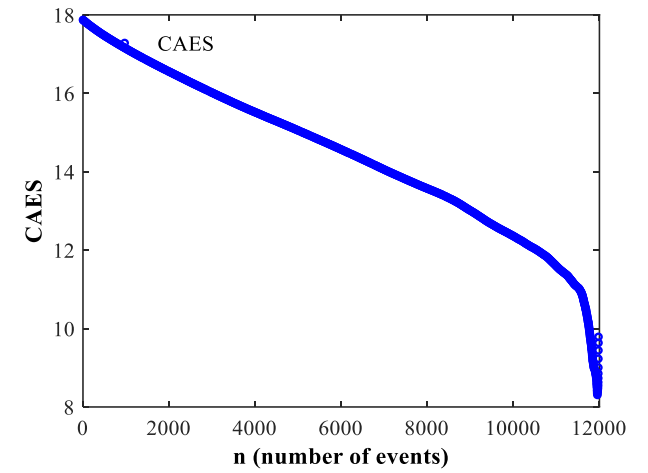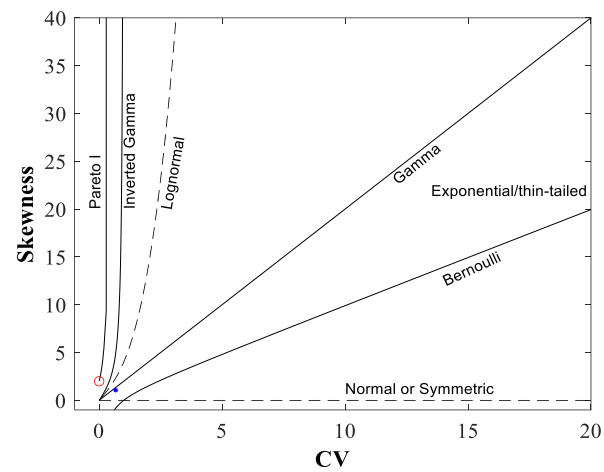
(i) MEF

(ii) CP

(iii) CAES

(iv) DMR plot

Figure 5.12 MEF, CP, CAES, and DMR plot for Weibull distribution $(\alpha = 1.491, \beta = 52.72)$ with sample size $n = 12000$
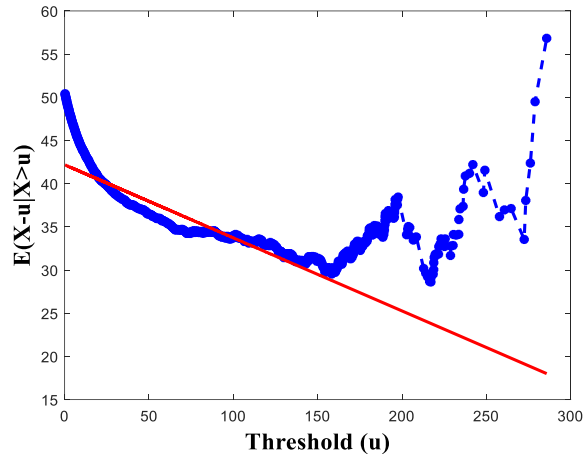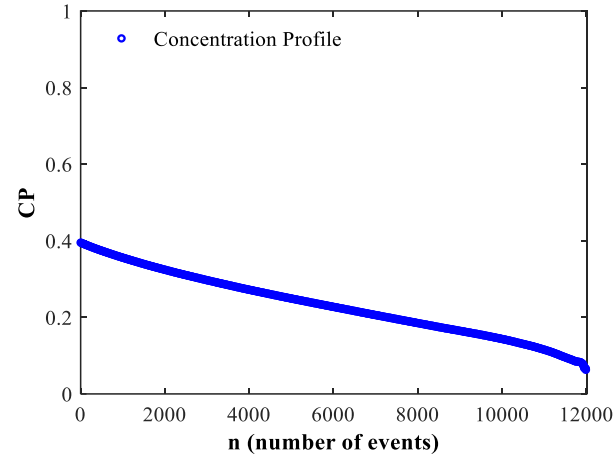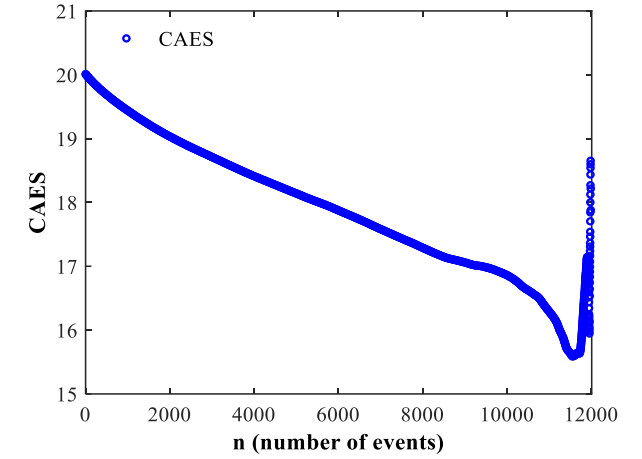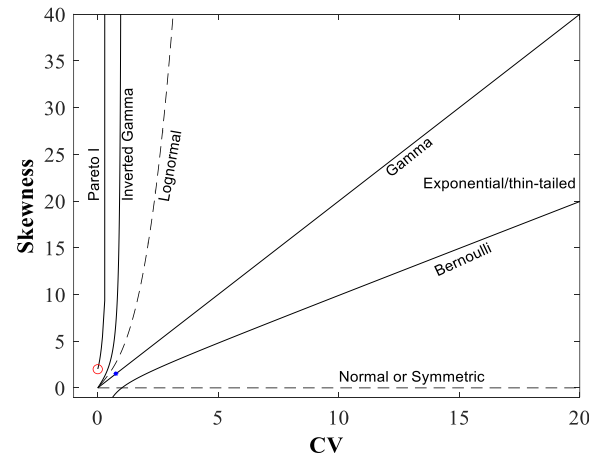
(i) MEF

(ii) CP

(iii) CAES

(iv) DMR plot

Figure 5.13 MEF, CP, CAES, and DMR plot for Gamma distribution $(\alpha = 1.8,\ \beta = 28.18)$ with sample size $n = 12000$.
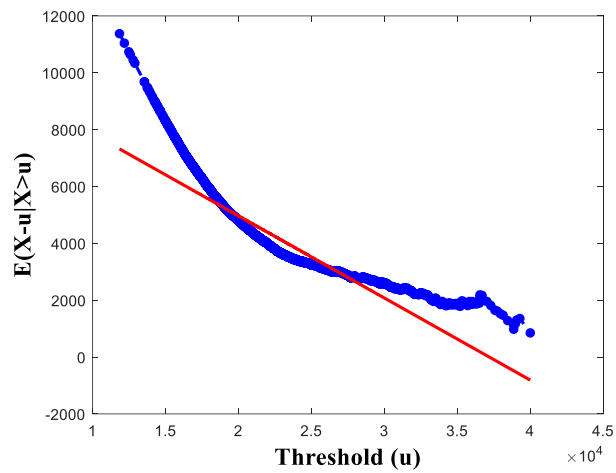
(i) MEF
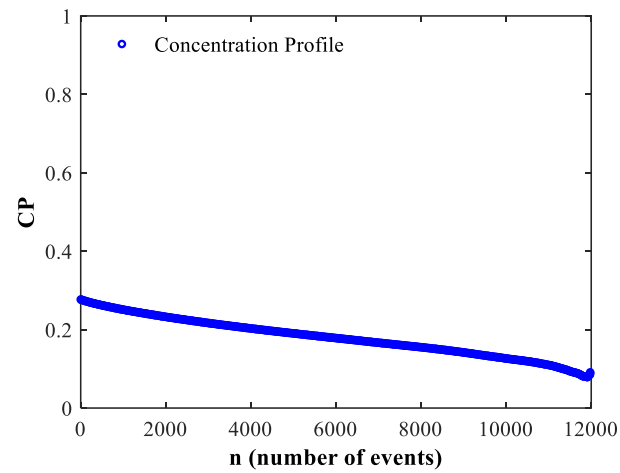
(ii) CP
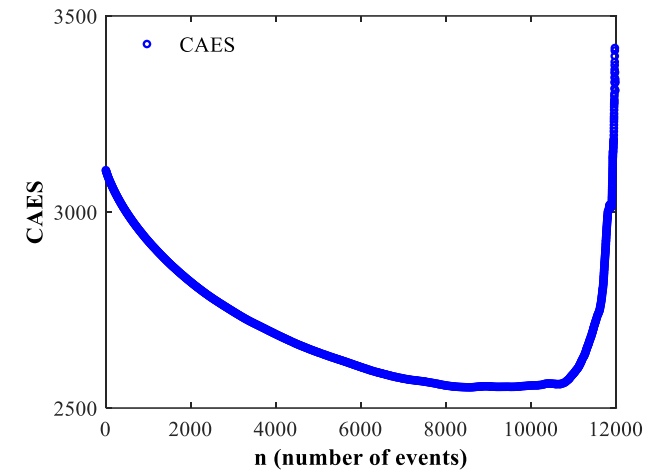
(iii) CAES
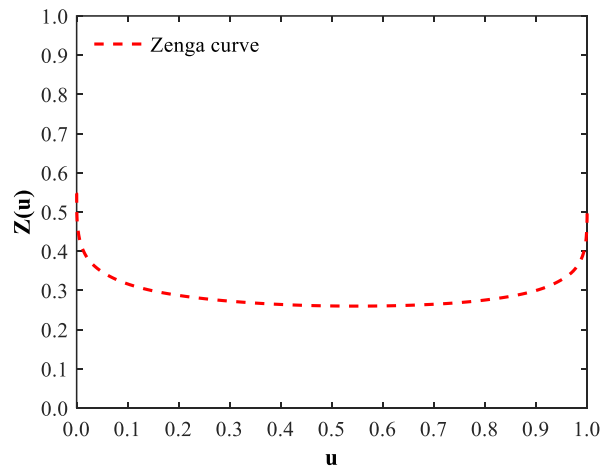
(iv) Zenga plot
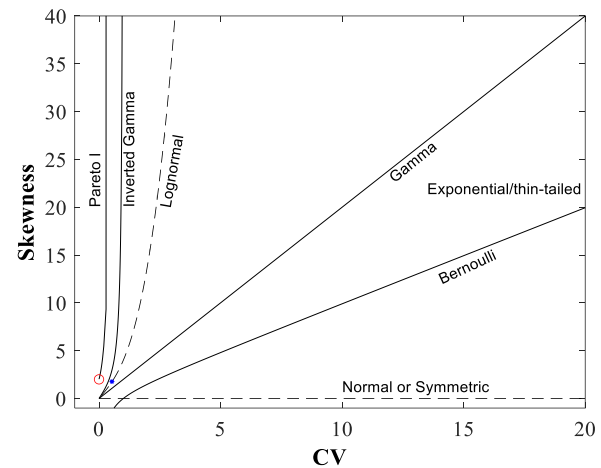
(v) DMR plot

Figure 5.14 MEF, CP, CAES, Zenga, and DMR plot for Lognormal distribution $(\alpha = 0.5,\ \beta = 9.6792)$ with sample size $n = 12000$

A constant Zenga curve close to 0.27 suggested a small value of the shape parameter, preferably lower than 1, confirming the findings of the CAES plot in the previous step. Similar observations were obtained for all 20 cases considered for each distribution in the hyper-exponential distribution category.

Overall, the simulation study revealed that the modified DSS was able to identify and characterize the different distributions into their respective classes. It should be noted there were some misidentified cases based on the CI approach of MEF usually occuring at smaller sample lengths (like n=500, 1000). For such cases, we recommend visual inspection of the MEF plots and the corresponding CP and CAES profiles. However, overall there was no ambiguity in characterizing the tails of different probability distributions into their respective classes based on the proposed DSS while identifying the appropriate class as well as confirming the class with confirmatory tests (i.e., class identified based on any test was never proved wrong in the confirmatory test). A significant reason behind this flawless performance of the proposed DSS was due to the inclusion of the state-of-the-art graphical method (e.g., CP, CAES, Zenga, DMR etc.) within the DSS framework. The CP and CAES are based on the sequence of the truncated Gini index. The Lorenz curve helps to visualize inherent heterogeneity in the datasets while the Gini index quantifies it. However, these methods are insufficient to completely characterize the right tail of the distribution (Fontanari et al., 2018a). Hence, the CP and CAES based on the truncated Gini index form more sophisticated methods which can provide a unique characterization of different distributions. These methods ease the identification of the distributions like Pareto (i.e., class B\A), LN, and Exponential. It also suggests the criterion that allows the discrimination of samples generated from LN and class C distributions. Followed by the CP and CAES, the tools like log-log plot, MEF plot, Zenga plot, and DMR (considered in the proposed DSS) allow the comparison of various distributions, simultaneously. However, combining various graphical methods in the correct order is essential to ensure the effective performance of any DSS. The findings of the simulation study revealed that the order or the combination of various graphical methods considered while framing the proposed DSS gives a reasonable degree of confidence when selecting the appropriate class of distribution. Overall, the simulation study revealed that the modified DSS is a potential tool for tail characterization.

5.7   *Minimum Sample length required for optimal performance of proposed DSS*

The conventional DSS, as well as the proposed DSS, utilizes graphical approaches to characterize the datasets into various classes of distributions. These graphical approaches may be sensitive to the sample size of the datasets considered for plotting. Thus, it becomes necessary to investigate the sensitivity of the DSS to the various sample sizes. To achieve this, sample sizes (n) of 500, 1000, 5000, 10000, 12000, 15000, 20000, and 25000 are considered for the generation of datasets from various distributions belonging to the aforementioned classes in section 5.6. Following this, the proposed DSS was implemented on each dataset to examine whether the DSS could perform the characterization correctly or not. Table 5.2 presents the percentage of success in characterization performed by the DSS. It can be noted that the characterization of distributions in various classes can be efficiently achieved

for datasets having a sample size greater than 1000. Hence, it is recommended that the proposed DSS can characterize the tails of the datasets having a sample size greater than 1000, which is generally the case for hydro-meteorological variables such as daily precipitation.

Table 5.2 Percentage of samples correctly identified by the DSS for each distribution considered for this study

Class of Distribution	Name of the Probability distribution	No. of parameter set considered for data generation	Sample Size (n)							
			500	1000	5000	10000	12000	15000	20000	25000
			Percentage of success in characterization (%)							
Class B\A	Pareto	30	100	100	100	100	100	100	100	100
LN	Lognormal ($\alpha > 1$)	30	80	93.33	100	100	100	100	100	100
Class D\C	Weibull ($\alpha < 1$)	30	80	86.66	100	100	100	100	100	100
Class D\C	Gamma ($\alpha < 1$)	30	83.33	86.66	96.66	100	100	100	100	100
Class C\B	Inverse Gamma	20	80	85	95	100	100	100	100	100
Class C\B	Log Pearson 3	20	70	90	95	100	100	100	100	100
E	Exponential	20	100	100	100	100	100	100	100	100
Hyper-exponential (outside Class E)	Weibull ($\alpha > 1$)	20	80	85	100	100	100	100	100	100
Hyper-exponential (outside Class E)	Lognormal ($\alpha < 1$)	20	75	80	90	100	100	100	100	100
Hyper-exponential (outside Class E)	Gamma ($\alpha > 1$)	20	75	100	100	100	100	100	100	100
Hyper-exponential (outside Class E)	Normal	20	100	100	100	100	100	100	100	100

5.8 Real-world application of Proposed DSS

In this section, the proposed DSS is applied to identify an appropriate class of distributions that describe the tails of daily precipitation data over India.

5.8.1 Description of the study area and data used

The study area considered for application of the proposed DSS is the entire India. The geographic location of India, along with the 34 contiguous meteorological subdivisions, is already shown in Figure 3.1 in Chapter 3. India is the largest South Asian country with a wide variety of climatic regions extending from low-precipitation arid regions to heavy precipitation receiving regions. The climate of India is influenced by various geographical and relief features like the Himalayas in the north, Thar Desert and Arabian Sea in the west, the Bay of Bengal in the east, Western Ghats in the southwest, and the Indian Ocean in the south. The study area experiences a wide range of variations in the characteristics of precipitation extremes. Thus, it is necessary to study the tail behavior of precipitation over India. In the present study, the daily gridded precipitation records (in mm per day)having a resolution of 0.25° were collected from the India Meteorological Department (IMD), Pune (Pai et al., 2014). Nearly 4801grids were considered for the analysis. The selection of the grids was done based on (i) the availability of continuous daily precipitation data for over 119 years (1901-2019) and (ii) the presence of at least 3000 nonzero daily values at each grid (Nerantzaki and Papalexiou, 2019). The sample size of non-zero daily precipitation data at the grids considered in this study ranges from 3015 to 27000.

5.8.2 DSS based characterization of tails of probability distributions of daily precipitation over India

The simulation study in section 5.6 established the efficacy of the proposed DSS in characterizing the tails of probability distributions belonging to various classes. This section demonstrates the effectiveness of the DSS through its application to daily precipitation data over India. It was envisaged to characterize the precipitation data at 4801 grids into various classes, namely, Class B\A, Class C\B, LN (the limiting case between Class C\B and D\C), Class D\C (Subexponential distributions), Class E (Exponential type distributions) and Hyper- exponential family (outside Class E) based on their tail behavior. The daily precipitation data at each grid was subjected to the procedure mentioned in the flow diagram (Figure 5.3). The MEF plots were first constructed for each grid over India to investigate whether the tail of the observed daily precipitation data at the grid is subexponential, exponential, or hyper-exponential. Based on Nerantzaki and Papalexiou (2019) approach, the estimate of the MEF slope was tested against the null hypothesis of zero slope indicating exponential tail by considering a 90% confidence interval. We found that the hypothesis of the presence of an exponential tail is rejected for nearly 4695 (97.79%) of grids over India at a 90% confidence interval (refer to Figure 5.15).
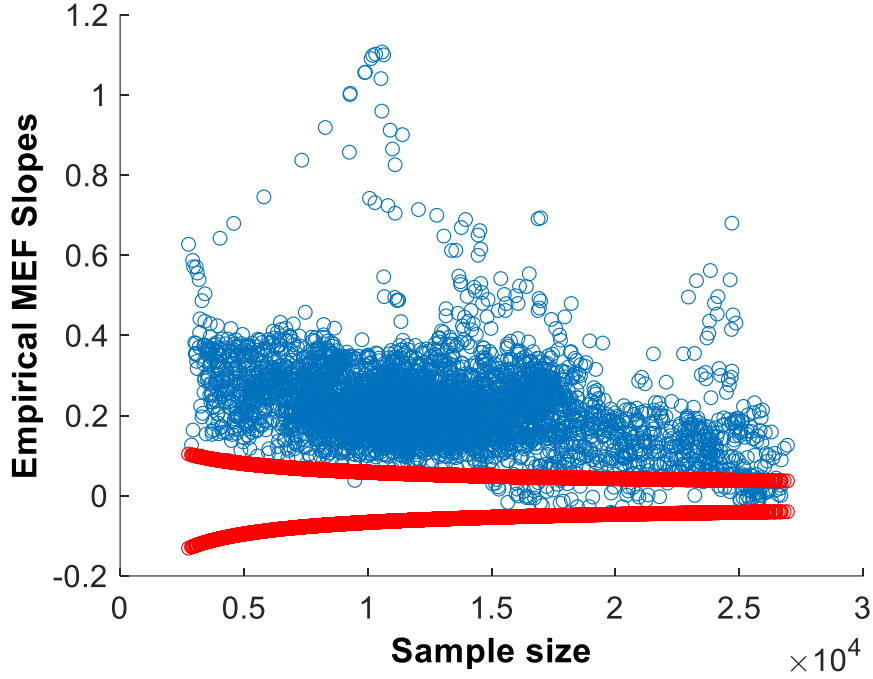
Figure 5.15 Plots of empirical MEF slope *vs.* the sample size *n,* along with the 90% CI for the exponential tails obtained from 4801 daily precipitation records over India

Furthermore, analysis based on a one-tailed test revealed that all the grids where the null hypotheses of the exponential tail are rejected, possess sub-exponential tails at a 10% significance level as the MEF slope at these grids lies above the upper Confidence Interval. So overall, out of 4,801 grids over India, 106 (i.e., 2.18%) grids indicated the presence of exponential tails, while 4695 (97.79 %) grids indicated the existence of subexponential tails. No grid was observed to possess a negative MEF slope suggesting the inappropriateness of hyper-exponential type distributions for fitting daily precipitation tails over India. Subsequently, the grids that showed the exponential tails based on the MEF approach were subjected to the CP and CAES approaches to validate the existence of exponential tails. Those approaches confirmed that the daily precipitation data at all 106 grids possessed exponential tails, and thus, the DSS suggested using a distribution of class E for these grids. Based on the analysis, the distributions from class E, such as exponential and Gamma ($\alpha = 1$), might be used to fit the precipitation data recorded at 106 grids over India. For brevity, the DSS-based characterization of the daily precipitation recorded at a grid (latitude 27° N and longitude 94° E) is shown in Figure 5.16.

The grids at which the MEF approach suggested the subexponential class of distribution, CP and CAES plots were used to further identify the subclasses (e.g., Pareto, lognormal, and other heavy tails, which were categorized as unidentified heavy tails). It was observed that none of the CP plots at any grid over India exhibited a constant behavior, i.e., horizontal line. This indicates that there is an absence of a Paretian tail in the daily precipitation data. Based on the CP and CAES plots, 961 grids (approximately 20.02% of 4801 grids) showed lognormal distribution ($\alpha > 1$) type tails. This is further

confirmed through the Zenga and DMR plots. The results are illustrated for a typical grid having a latitude 33° N and longitude 78.5° E in Figure 5.17. For this grid, the empirical CP exhibited the following features: it starts around the point $G(0) \approx 0.6825$ (which is > 0.5); it decays towards zero; and has a change in the slope being slightly convex at the beginning and concave at the end. The visual interpretation of CP helps to eliminate the Pareto class of distribution and indicate a lognormal type distribution. The diverging CAES also points to the presence of lognormally distributed data and provides some additional information that the shape parameter of the lognormal distribution would be greater than 0.7 (in this case, greater than 1). Figures 5.17 (d) and 5.17 (e) show Zenga and DMR plots which are part of the confirmatory test. The Zenga plot showed a constant behavior except at extremities, and the couple point $(\hat{\gamma}_2, \hat{\gamma}_3)$ in the DMR plot was found to lie in the lognormal zone, close to the lognormal curve.

After identifying the limiting cases of lognormal, the discrimination between classes C\B and D\C for the remaining grids was needed. It should be noted that the CP for any heavy-tailed distribution belonging to D\C classes tends to have $G(0) > 0.5$. Few distributions like Gamma $(\alpha < 1)$ and Weibull $(\alpha < 1)$ can appear similar on a CP plot for a considerable data length; hence to avoid ambiguity, they are commonly categorized under the category of unidentified heavy tails. There were 3734 (77.77%) grids out of 4801, which were included in the category of unidentified heavy tails. After this, a log-log plot was employed to further categorize them following the flow diagram of DSS. A straight-line log-log plot indicates that the data belong to class C\B instead of class D\C. Out of 3734 grids, 73 (1.52% grids out of 4801 grids) and 3661 (76.25% grids out of 4801 grids) grids belonged to classes C\B and D\C, respectively. Subsequently, the MS plot (Hill ratio plot) was used as a confirmation test for the grids falling in class C\B (D\C).
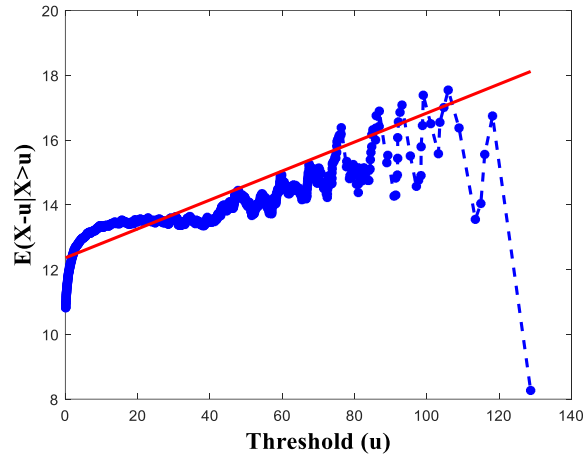
(i) MEF
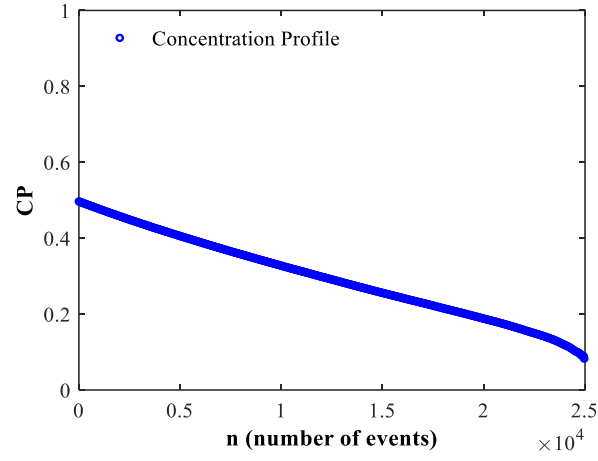
(ii) CP
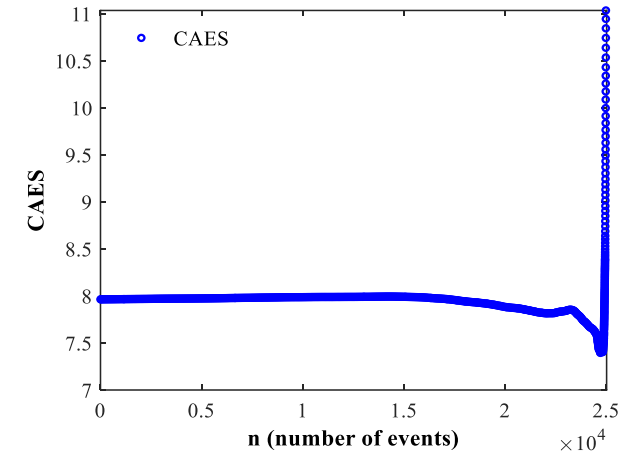
(iii) CAES
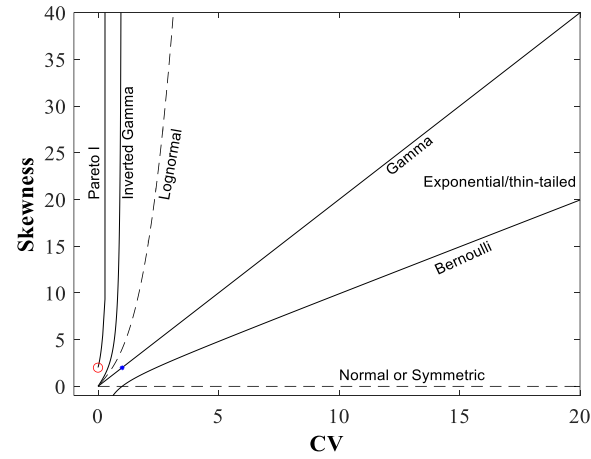
(iv) DMR plot

Figure 5.16   DSS-based characterization of daily precipitation recorded at a typical grid (latitude 27° N and longitude 94° E) corresponding to class E
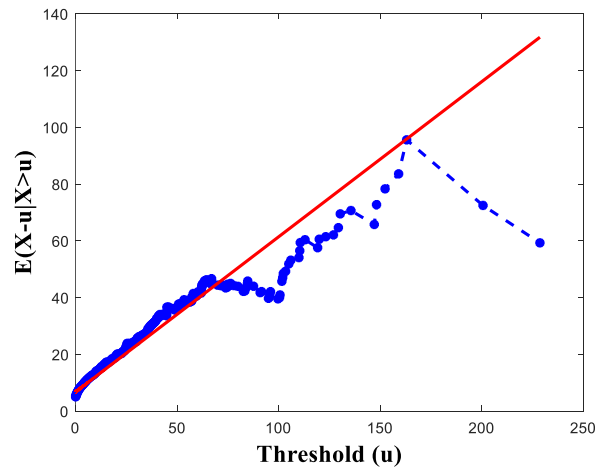
(i) MEF
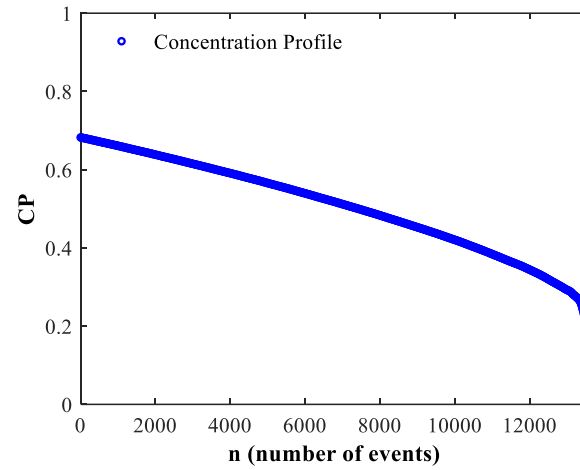
(ii) CP
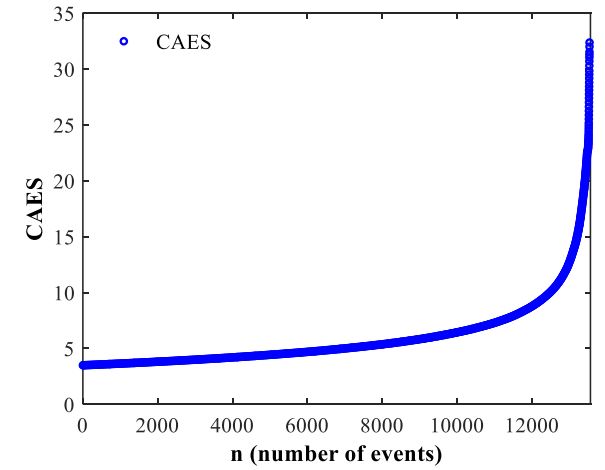
(iii) CAES
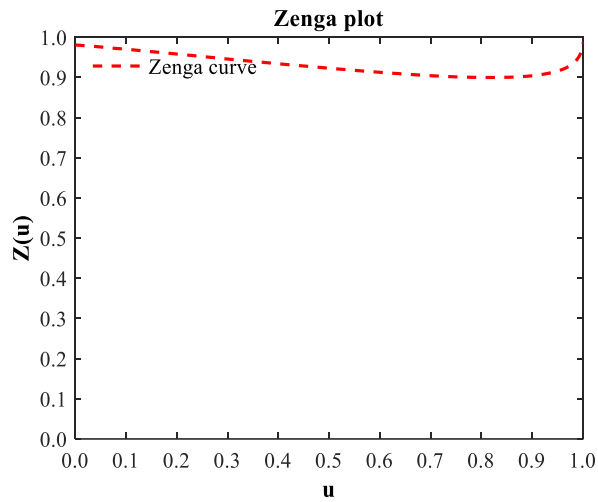
(iv) Zenga plot
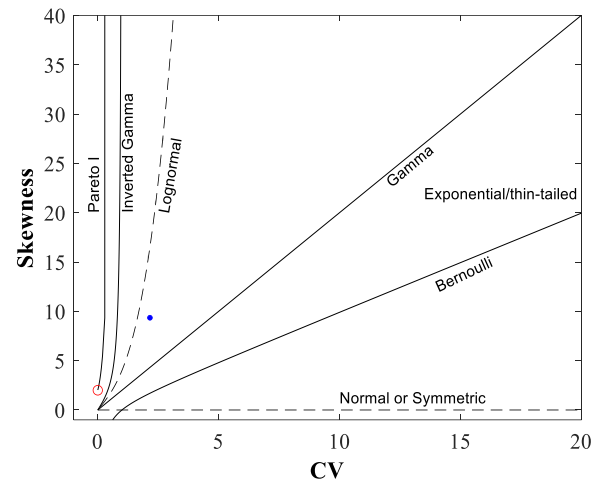
(v) DMR plot

Figure 5.17 DSS-based characterization of daily precipitation recorded at a typical grid (latitude 33° N and longitude 78.5° E) corresponding to *LN*

Figure 5.18 and 5.19 illustrate the characterization of two typical grids corresponding to latitude 37.25° N, longitude 74.5° E, and latitude 20.75° N, longitude 82.75° E representing the classes C\B and D\C, respectively. It can be noted from the figures that the log-log plot in Figure 5.18 (iv) is more close to a straight line than the plot in Figure 5.19 (iv), which does not follow a straight line. Therefore, the DSS suggests the use of a distribution of Class C\B for the grid shown in Figure 5.18, while it suggests the use of a distribution of Class D\C for the grid shown in Figure 5.19. Finally, the confirmation of the classes for both grids was done based on the MS plot (Figure 5.19 (v)) and Hill ratio plot (Figure 5.19 (v)). The MS plot (see Figure 5.18 (v)) illustrates that the plot does not converge to zero for values of p greater or equal to 3, which means that only the 1st and 2nd raw moments exist for the data while the other moments do not exist. This confirms that the data belongs to class C\B (i.e., regularly varying distributions). Based on the analysis, the distributions from class C\B, such as Fréchet, HIB, IG, and LP3 might be used to model the precipitation data recorded at 73 grids over India. Figure 5.19 (v), showing the Hill ratio plot, indicates the convergence of the plot towards zero, which confirms that the underlying distribution for the data at the grid comes from a subexponential class (i.e., class D\C). Consequently, the distributions from class D\C, such as Halphen A, Gumbel, Gamma, or Pearson type III might be used to model the precipitation data recorded at 3661 grids over India. The DMR plot in Figure 5.19 (vi) shows that the gamma distribution can be a suitable candidate for this particular grid. However, the choice of suitable candidate distribution within a class can be made using classical tests and criteria such as the Anderson darling test, Akaike Information Criterion, Bayesian Information Criterion, or New Model selection test (NMST) (Akaike, 1974; Schwarz, 1978; Ehsanzadeh, 2010; Haddad and Rahman, 2011; Panahi, 2016; Nassa et al., 2021).

Figure 5.20 illustrates the spatial distribution of classes providing the best fit to the non-zero daily precipitation data over India based on the proposed DSS. Overall, it can be inferred from the analysis that the majority of the precipitation data recorded at various grids over India possess heavy tails (i.e., 4695 out of 4801). With the dominance of heavy tails all over India, few pockets comprising lighter tails (i.e., 106 grids out of 4801) were observedin some parts of Kerala, the northeast region (Arunachal Pradesh, Naga, Mani and Tripura), and along the western coastal plain of the Indian Peninsula (i.e., in Coastal Karnataka, Konkan and Goa). Gupta and Chavan (2022) analyzed the tails of daily precipitation based on the concept of obesity index and noticed similar behavior of tails along the western coastal plain, Kerala, and northeast region.

(i) MEF

(ii) CP

(iii) CAES

(iv) Log-log plot

(v) MS plot
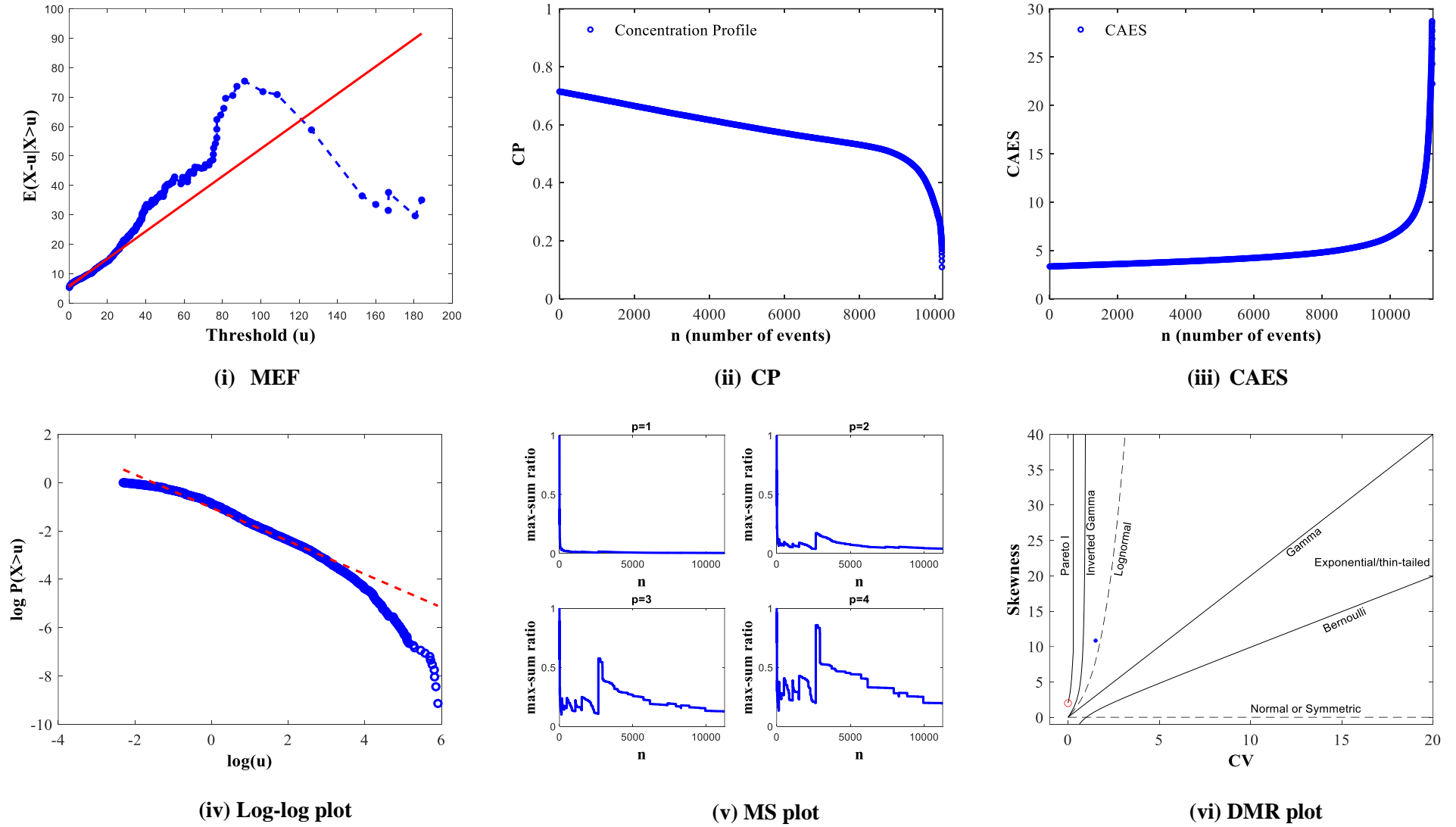
(vi) DMR plot

Figure 5.18 DSS-based characterization of daily precipitation recorded at a typical grid (latitude 37.25° N and longitude 74.5° E) corresponding to class C\B
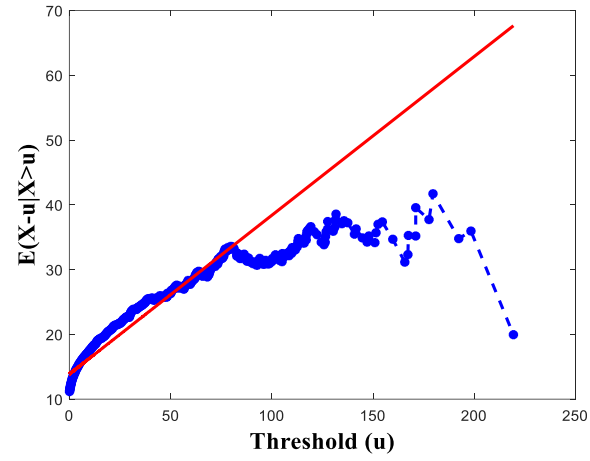
(i) MEF
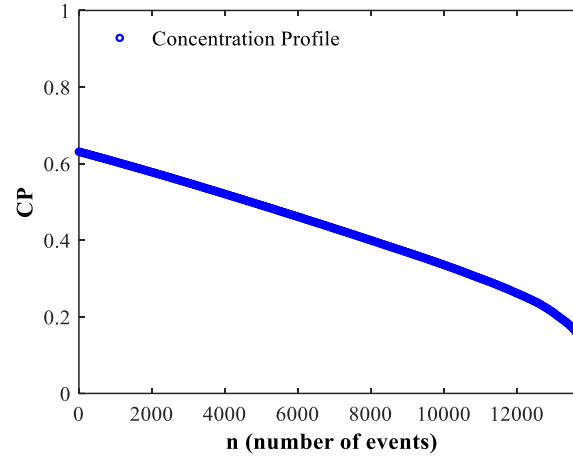
(ii) CP
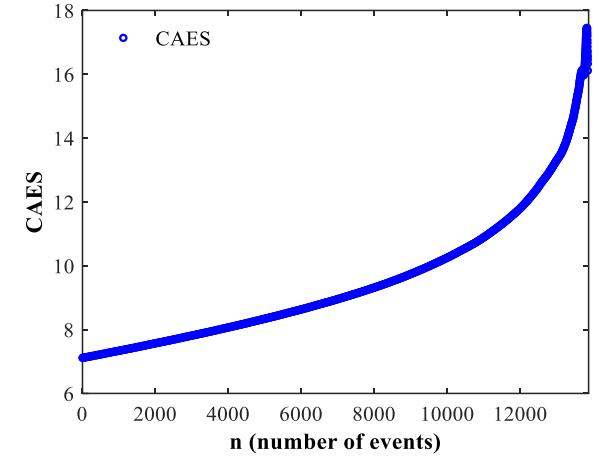
(iii) CAES
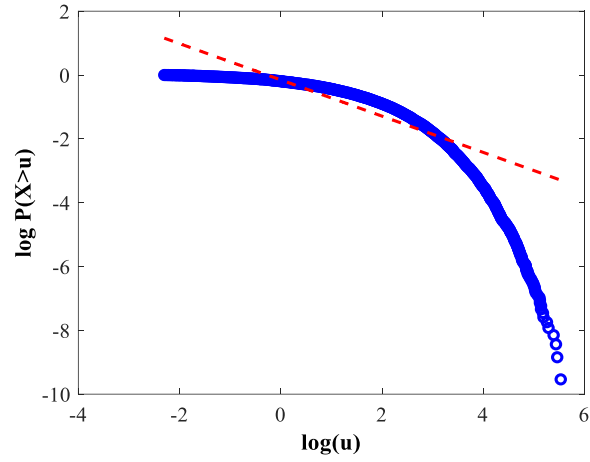
(iv) Log-log plot
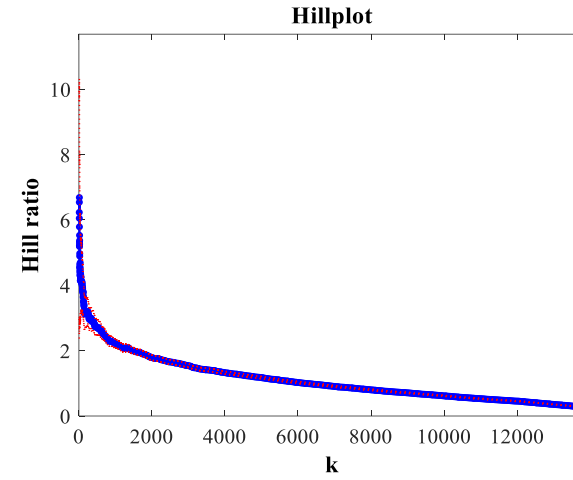
(v) Hill ratio plot
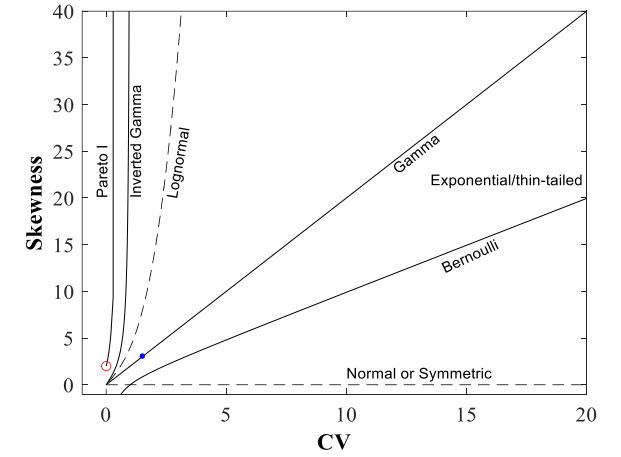
(vi) DMR plot

Figure 5.19 DSS-based characterization of daily precipitation recorded at a typical grid (latitude 20.75° N and longitude 82.75° E) corresponding to class D\C
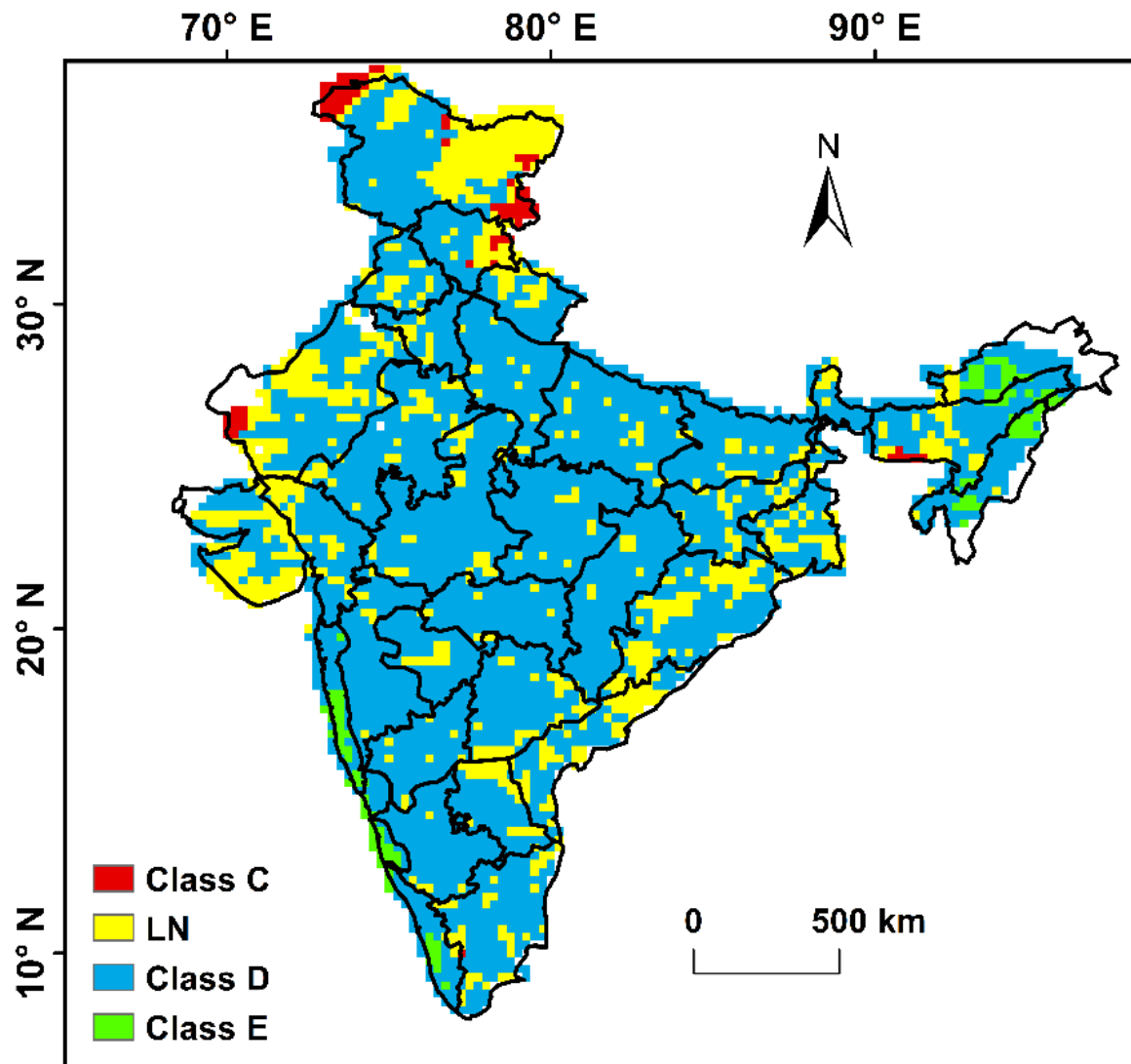
Figure 5.20 Spatial distribution of classes providing the best fit to on-zero daily precipitation over India based on the DSS

Heavy tails belonging to the lognormal class are found to be spread all over India, with maximum presence in regions like Saurashtra Kutch and Diu, Gujarat, West, and East Rajasthan, the western part of Jammu and Kashmir, Punjab, Uttaranchal, Bihar, Gangetic West Bengal, Jharkhand, Orrisa, and Coastal Andhra Pradesh. At the same time, the heaviest tails belonging to class C\B are found at very few locations in Jammu and Kashmir, Himachal, and Rajasthan. Previous studies over India demonstrated that the frequency of extreme precipitation over India is increasing in the regions of heavy tails predominantly. (e.g., Goswami et al., 2006; Ajayamohan and Rao, 2008; Dash et al., 2009; Guhathakurta et al., 2011; Vittal et al., 2013; Mishra et al., 2014; Krishnan et al., 2016; Roxy et al., 2017; Sarkar and Maity, 2020; Yaduvanshi et al., 2021; Gupta and Chavan, 2022). The increased frequency of extreme precipitation over India can be attributed to the abrupt global change of the climatic system resulting from the climate shift of the 1970s in various climate phenomena like the Arctic Oscillation, East Asian summer monsoon, East Asian winter monsoon, North Atlantic

Oscillation, El Niño–Southern Oscillation, Aleutian low, Pacific decadal oscillation, Western Pacific subtropical high, and Indian summer monsoon rainfall (Biondi et al., 2001; Chowdary et al., 2006; Zhou et al., 2009; Vittal et al., 2013; O'Kane et al., 2014; Sabeerali et al., 2012; Sahana et al., 2015; Dai et al., 2018).

It should be noted that the presence of a heavy-tailed class within a region does not necessarily mean that the region would receive high-magnitude precipitation events. For instance, the western coastal plain of the Indian Peninsula (like Konkan and Goa, Coastal Karnataka) tends to receive the highest amounts of precipitation in India, but as the amounts are consistently high, the daily precipitation tails are observed to be lighter. However, in the arid and semiarid parts like Rajasthan, Gujarat, Saurashtra, Kutch, and Diu, daily precipitation exhibits heavier tails as the extreme precipitation events are much larger than the average. Past studies have also shown that heavy tails are primarily found in regions where multiple types of synoptic systems occur, while the exponential distribution is generally seen in the areas dominated by a single type of weather event (Kozubowski et al., 2009; Cavanaugh et al., 2015). The characterization of most of the grids over India into the heavy-tailed classes (i.e., class C\B, D\C) and heavy-tailed distribution LN points towards the frequent occurrence of extreme precipitation events in most parts of the country. Similar findings were observed by Nerantzaki and Papalexiou (2019) when they analyzed global daily precipitation data based on algorithmic MEF. They found heavier tails (based on the exceedance probability of the observed MEF slopes) over the western U.S., south-western Canada, India, and central China. The high frequency of extreme precipitation events can cause severe damage to water treatment plants and sewage networks and play a key role in waterborne disease outbreaks, thereby impacting public health (Curriero et al., 2001). It can also affect the agricultural productivity of the country by damaging crops (Rosenzweig et al., 2002). Hence, such events are needed to be identified and modelled correctly.

Topography or elevation also plays a critical role in shaping the local climate and weather patterns as it impacts temperature, moisture content, and atmospheric stability at different altitudes. These variations have a significant effect on the occurrence of extreme weather phenomena such as heavy rainfall, intense storms, and extreme temperatures. In regions with mountains or high elevations, cooler temperatures at higher altitudes lead to increased atmospheric stability and a higher probability of precipitation. This results in a greater likelihood of intense rainfall or snowfall, which can lead to flooding or avalanches. This observation implies a potential relationship between the heaviness of precipitation tails, elevation, and the Köppen-Geiger climate classification system. The Köppen-Geiger system utilizes various variables, including temperature, to classify the climate of specific regions (Köppen and Geiger, 1923; Peel et al., 2007; Rubel and Kottek, 2010; Beck et al., 2018; Hobbi et al., 2022).

Based on the above-mentioned observations regarding the spatial variation of tail heaviness, it was observed that the spatial distribution of the tail heaviness is closely related to the climate zones and

topography of the country. Figure S.82 in the supplementary material of Gupta and Chavan (2023b) shows different climate zones based on the Köppen-Geiger classification (Rubel and Kottek, 2010). The regions where heavy (class D) and heaviest (LN and class C) tail are observed belongs to the climate zone like Arid (Bwh, BSh), Polar (ET), and Tropical (Aw, As), based on the Köppen-Geiger climate classification over India. The grids associated with climate type Cwb and Am possess less heavy tails (like exponential). Earlier studies like Gupta and Chavan (2022) and Papalexiou et al. (2018) have also reported similar findings of tail behavior in Polar (E), Arid (B), and Tropical (A) regions. Rajulpati et al. (2020) examined 5 global gridded precipitation datasets and found that products like Climate Prediction Center, Multi-Source Weighted Ensemble Precipitation, and Water and Global Change (WATCH) Forcing Data–ERA-Interim indicate medium tails, while the Climate Forecast System Reanalysis dataset indicates heavy tails over India. They also observed heavy tails in hot, arid climates, medium to light tails in temperate climates, and lighter tails in coastal regions worldwide. Overall, they observed medium to heavy tail over Asia (which includes India) based on the tail index. Their results are in close correspondence with our findings. Further, we also noticed that the grids associated with climate-type ET and BWh have the heaviest-tails. Papalexiou et al. (2018) found that some of the heaviest tails are seen in the mountainous and arid regions, while Hobbi et al. (2021) in their study found the heaviest tails for Climate type E (Polar); our results agree with both studies as most of the ET zone in India comprises hilly areas. However, for more accurate insights into the spatial variation of tails and their relationship with elevation in specific regions detailed localized studies and analysis is important.

5.9   Summary and Conclusions

The present study proposes a comprehensive Decision Support System (DSS) for the characterization of tails of probability distributions of daily precipitation data over India. The DSS proposes an algorithmic way of utilizing advanced graphical methods in a particular order to ensure the efficient identification of classes of distributions based on tail behavior. This study shows that the proposed DSS can characterize the tails of probability distributions into various classes by alleviating a number of the limitations associated with the conventional DSS. The proposed DSS is successful in discerning the tails of probability distributions belonging to the hyper-exponential family, exponential family, subexponential family, and Lognormal distribution. The following conclusions are drawn from this study.

(i)   The robustness of the proposed DSS over the conventional approaches is established through a simulation study where datasets having various sample sizes from representative probability distributions belonging to various classes (e.g., D, C\B, B\A, E, etc.) and limiting case LN are generated, and reassigning them successfully into their respective classes.

(ii)   The proposed DSS utilizes the ability of Concentration profile and Concentration adjusted expected shortfall to discriminate between Pareto distribution (i.e., class B\A) and class D

(which include Lognormal distribution as well as distributions from classes C\B and D\C). Further, the DSS assists distinction between the distributions from class D through the use of a log-log plot, MEF plot, Zenga plot, MS plot, and DMR plot in addition to CP and CAES in an appropriate order.

(iii) Based on the proposed DSS, the daily precipitation data over India was successfully characterized into 4categories, namely class C\B, D\C, E, and limiting case LN. Results indicate that out of 4801 grids, 63 (1.31%), 3671 (76.46%), 106 (2.21%), and 961 (20.01%) grids belong to classes C\B, D\C, E, and LN distribution, respectively. Overall, around 98% of the daily precipitation records are found to be exhibiting subexponential tails. The spatial information on these classes of distributions over India could be used as a priori information for regional studies to identify suitable probability distributions.

(iv) Heavy tails belonging to the lognormal class are found to be extending all over India, with maximum presence in regions like Jammu and Kashmir, Saurashtra Kutch and Diu, Gujarat, West and East Rajasthan, Gangetic West Bengal, Punjab, Bihar, West Bengal, Uttaranchal, Jharkhand, Orrisa, and Coastal Andhra Pradesh. While the heaviest tails belonging to Class C\B are found at very few locations in Jammu and Kashmir, Himachal, and Rajasthan.

In summary, the correct identification of the distribution families describing the precipitation tails is necessary for the reliable estimation of quantiles corresponding to different return periods. An inappropriate model selection due to selecting a wrong class of distributions might lead to the underestimation or overestimation of the quantile estimates, thereby increasing the socio-economic risk. The proposed DSS can be considered as an effective and easy-to-use tool for exploratory analysis and identifying the classes of distributions that provide the best fit to a dataset, especially the right tail. However, It may be noted that the application of the present DSS can be a little strenuous when a large dataset is considered due to the visual assessments required for categorization, especially between Class C and Class D\C. There is a need for an algorithmic mechanism to discriminate between the distributions belonging to Classes C\B and D\C by developing the procedure for examining the linearity of the log-log plots based on estimated correlation coefficient values, similar to the Mean Excess Function (MEF) proposed by Nerantzaki and Papalexiou (2019). To achieve this, a simulation study needs to be performed involving the construction of the confidence interval representative of all the distributions from Class C. This can automate the analysis. Further research related to the development of such algorithmic mechanisms is underway. These advancements are expected to address the limitations of the present DSS and provide more automated and efficient methods for model selection and decision-making for practicing engineers.

Overall, even in its current form, the decision support system is believed to be valuable in assisting engineers with appropriate model selection and making informed decisions.

ASSESSMENT OF EMBEDDED RISK IN PRECIPITATION TAILS OVER INDIA THROUGH CONCENTRATION PROFILES: A MULTI-MODEL ASSESSMENT FROM CMIP6 EXPERIMENTS

6.1 Overview

In a country like India, changingg climate has led to increased occurrences of extreme precipitation events causing extensive damage to humans and nature. Understanding of the altered frequencies and magnitude can be done based on the assessment of the tails of the probability distributions that can represent them both. Researchers have explored these variations by examining the tail heaviness of daily precipitation datasets using different scalar and graphical methods. However, little is known about the actual risk associated with these extremes, especially among different climate zones of India. Tail risk (low-probability extreme events) technically refers to the risk associated with both the left and right tails. People working in the field of finance and economy are mostly concerned with losses (the left tail), while hydrologists or climatologists are generally concerned with extremes belonging to the right tail. The utilization of unconventional tools based on the concentration measures of distributions to quantify the tail risk of the precipitation datasets has not been done till now. Hence, this chapter shows the utility of a simple and novel risk assessment technique known as a Concentration Map (CM) that provides a single concise value after analyzing the riskiness of the time series. Concentration Map uses inferences from the Concentration Profile (CP), which is a novel way of characterizing and estimating the tail variability, based on the concentration measures (like the Gini Index).

This study presents a comprehensive analysis in which we quantify the tail risk associated with the daily precipitation extremes for the past and future across India. Risk embedded into the tails was evaluated for around 4801 gridded station-based datasets for the historical time period, while the simulations from 16 General Circulation Models (GCMs) participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6) under four Shared Socioeconomic Pathway (SSPs), namely, SSP126, SSP245, SSP370, and SSP585 are considered for future at those same stations. It is noteworthy that there remains a dearth of attempts to assess and identify tail risks using innovative risk assessment methodologies for daily precipitation datasets across India, especially considering all four SSPs, namely SSP126, SSP245, SSP370, and SSP585 under the new CMIP6 scenarios.

Furthermore, the analysis is extended to conduct a regional scale investigation on how the risk varies according to different meteorological subdivisions as well as for Indian Climate zones. Explorations related to the tail risk variation within different climate zone is done to identify the climate type more prone to extremes. The most widely used scheme to define the climate zone is the Köppen-Geiger (KG) classification, which groups climate across the globe into several subtypes using multiple variables based on precipitation and temperature. Studies that apply the KG classification serve multiple purposes, like illustrating the geographical spread of climate types (or zones) and examining the characteristics and changes occurring within specific climate types in relation to extreme events. While numerous studies focusing on assessing extreme events and their tail behavior across various climate zones are present, there is a notable lack of discussion regarding the associated risks linked with these extreme events.

6.2   *Theoretical Background on Concentration Map (CM)*

The concept of the Concentration Map was proposed based on the Concentration Profile, which is itself based on the utilization of concentration measures. The CP is a quick heuristic that is helpful in not only describing the variability within the parametric family of distributions but also in identifying the distribution that provides the best fit to the dataset, especially to its right tail (large extreme events). Concentration Maps overcome limitations associated with traditional risk management measures (Wietzke et al., 2020). The major advantage associated with the CM is that it maps the major risk factors contained in CP into an easily readable plot, which can further be converted into a risk score. It is an easy and scale-free approach that can be used to compare the distribution or datasets with different scales and magnitudes in terms of their tail risk (Fontanari, 2019). A brief description of the Concentration Profile is already provided in section 5.4.2 in the chapter while the details of the Concentration Maps is provided below.

6.2.1   *Concentration Map (CM)*

The CM, initially introduced by Fontanari et al. (2018a), is a graphical way of assessing the risk embedded in a distribution by making use of the CP of a particular distribution. It is basically a two-dimensional map formed by extracting the relevant information from the Concentration Profile and combining it with utility or risk theory (Edwards, 1992). CM identifies the main risk factors contained in the CP and maps them into an easily interpretable plot using a risk/utility function approach. Every CP has a concise risk score attached to it, which eases their comparison in terms of their tail or global risk.

Given a CP $\{G(\lambda)\}_{\lambda \in [0,1)}$, the main risk drives r_1 and r_2 i.e., the quantities that summarize most of the risk, are identified. The risk driver r_1 is the starting value of the CP, i.e., $r_1 = G(0)$ and the risk driver r_2 is given by the difference between the risk driver r_1 and $G(\lambda)$, with $\lambda \to 1$ (i.e.,

$r_2 = \lim_{\lambda \to 1} |G(0) - G(\lambda)|$). The rationale behind each of the risk drivers are that in case of r_1, a CP with higher $G(0)$ among different CPs with similar behavior is considered to be riskier (or fatter/heavier) and for risk driver r_2, smaller the gap between the initial and final values of the CP, larger the mass present in the tail, and the higher the probability of the extreme suggesting presence of a heavy tail. Once the risk drivers are identified from the CP, they are combined using a Cobb-Douglas risk function (Cobb and Douglas, 1928; Barucci and Fontana, 2003; Hassani, 2012) which provides a single concise measure of embedded risk, as given by Equation (6.1).

$$R(r_1, r_2) = r_1^a (1 - r_2)^b \text{ with } a, b \in R^+ \tag{6.1}$$

The above-mentioned risk function provides the risk values between 0 to 1 which not only summarizes the riskiness of a particular CP but also helps us in ranking different CPs based on their embedded risk. Cobb-Douglas risk function is a preferred choice for risk assessment as it allows us to (1) assess the relevant risk driver based on the value of both parameters, (2) represent output elasticities based on the estimated coefficients, (3) derive interpretable iso-risk curves and (4) substitute the risk driver with ease (Theriault and Serra, 2014; Fontanari et al. 2018a; Tong et al., 2019).

The approach of CM is a scale-free approach; hence, the comparison of multiple and different datasets having different scales and magnitudes is achieved with ease. However, the parameter values used to derive the CMs hold significant importance when assessing information about embedded risk for a particular distribution. Particularly, higher a value shows the relevance of the first risk driver which represents the global risk as it accounts for the entire distribution. While the higher value of the parameter b shows the relevance of the second risk driver which represents tails. The values of these parameters can be based on historical data or expert judgments. To have a better understanding about the role of the parameter values a simulation study is conducted and presented in the next section.

6.3 Simulation study to understand the nature and variation of risk based on the parameter values

To have a better understanding of the tail risk embedded in different distributions, we performed a simulation experiment where we generated datasets of various representative distributions from different classes of distribution (El Adlouni et al., 2008; Gupta and Chavan, 2023) and estimated their risk values based on Cobb-Douglas risk function for different combinations of a and b values. The datasets were generated for distributions like Weibull (Class D\C), Lognormal (limiting case between C and D), and Pareto (Class B). To have a better understanding of the tail risk associated with datasets of different distributions, we tried to keep the global risk value nearly equal. In order to do so the datasets were generated in such a manner that their parameter sets have similar or somewhat equal $G(0)$ or r_1. All datasets generated from different distributions were grouped into eight groups based on r_1. Details

related to the distribution, their parameter, and estimated risk function values for each group derived from the CM for a different combination of parameter a and b values are presented in Table 6.1. Figure 6.1 shows the concentration Maps generated based on the Cobb-Douglas risk function for four different combinations of parameter values a and b (for brevity only four CMs are shown). It is important to note that any color on the map that is equal to the one placed on the exponential cell or on the Pareto segment will share the same risk, due to the isorisk property of the Cobb-Douglas risk function (Edwards, 2013). This is an important feature of the Concentration Maps as a risk characterization tool.

Now, we know that for higher values of a, the relevance is of the first risk driver i.e., r_1. Since r_1 is the same for all the distributions belonging to one group, their global risk value is also similar at high a. For example at a=0.9 and b=0.1, all three distributions in different groups have similar risks. However, for higher b the relevance is of the second risk driver r_2 which represents tails and their risk. For example, in group 1 (see Table 6.1), as the value of parameter b increases risk is always higher for Pareto followed by Lognormal and Weibull. This finding aligns with the concept of nested classes of distribution, indicating that the distribution of Class B has heavier tails than that of Class C distributions, which in turn has heavier tails than LN. Furthermore, LN distributions have heavier tails than Class D and the distribution of Class D has heavier tails than Class E (El Adlouni et al., 2008). Similar observations are seen for other groups too. Hence, the simulation study findings collectively indicate that to highlight tail risk, the parameter selection should prioritize giving greater weight to parameter b. Consequently, guided by both the simulation results and expert recommendations, we will adopt the parameter values a=0.3 and b=0.7 for this study. This specific parameter combination has been deemed appropriate for evaluating tail risk across various datasets, and it will be useful in accurately assessing tail risk in daily precipitation datasets.

Table 6.1 Generated distribution details and their estimated risk function values from CM for various combinations of parameters *a* and *b*

Group No.	S. No.	Distribution Name	Parameter Values (α, β)	Risk driver r_1	Risk driver r_2	Risk Function Values								
						a=0.1, b=0.9	a=0.2, b=0.8	a=0.3, b=0.7	a=0.4, b=0.6	a=0.5, b=0.5	a=0.6, b=0.4	a=0.7, b=0.3	a=0.8, b=0.2	a=0.9, b=0.1
I	1	Weibull	(3, 10)	0.21	0.17	0.72	0.63	0.55	0.47	0.41	0.36	0.31	0.27	0.24
	2	Lognormal	(0.37,10)	0.21	0.14	0.75	0.65	0.56	0.49	0.42	0.37	0.32	0.28	0.24
	3	Pareto	(2.9, 10)	0.21	0.01	0.85	0.73	0.62	0.53	0.46	0.39	0.33	0.29	0.24
II	4	Weibull	(2.1, 10)	0.28	0.24	0.69	0.62	0.57	0.51	0.46	0.42	0.38	0.35	0.31
	5	Lognormal	(0.5, 10)	0.28	0.20	0.72	0.65	0.58	0.53	0.47	0.43	0.38	0.35	0.31
	6	Pareto	(2.5, 10)	0.28	0.03	0.86	0.76	0.67	0.59	0.52	0.46	0.41	0.36	0.32
III	7	Weibull	(1.8, 10)	0.32	0.27	0.67	0.62	0.57	0.53	0.48	0.45	0.41	0.38	0.35
	8	Lognormal	(0.58, 10)	0.32	0.20	0.73	0.66	0.61	0.55	0.51	0.46	0.42	0.38	0.35
	9	Pareto	(2,10)	0.32	0.06	0.85	0.76	0.68	0.61	0.55	0.50	0.44	0.40	0.36
IV	10	Weibull	(1.13, 10)	0.46	0.38	0.60	0.59	0.57	0.55	0.53	0.52	0.50	0.49	0.47
	11	Lognormal	(0.87, 10)	0.46	0.28	0.68	0.65	0.63	0.60	0.57	0.55	0.52	0.50	0.48
	12	Pareto	(1.4,10)	0.46	0.07	0.87	0.81	0.75	0.70	0.65	0.61	0.57	0.53	0.49
V	13	Weibull	(0.74,10)	0.61	0.48	0.53	0.54	0.55	0.56	0.57	0.58	0.59	0.60	0.61
	14	Lognormal	(1.2, 10)	0.61	0.35	0.65	0.64	0.64	0.64	0.63	0.63	0.62	0.62	0.61
	15	Pareto	(1.25,10)	0.61	0.13	0.84	0.81	0.78	0.76	0.73	0.70	0.68	0.65	0.63
VI	16	Weibull	(0.28, 10)	0.92	0.57	0.47	0.50	0.54	0.58	0.63	0.68	0.73	0.79	0.85
	17	Lognormal	(2.55, 10)	0.92	0.36	0.67	0.69	0.72	0.74	0.77	0.80	0.83	0.86	0.89
	18	Pareto	(0.95,10)	0.92	0.06	0.92	0.93	0.93	0.93	0.93	0.94	0.92	0.92	0.91
VII	19	Weibull	(0.2, 10)	0.97	0.42	0.61	0.64	0.67	0.71	0.75	0.79	0.83	0.87	0.92
	20	Lognormal	(3.5, 10)	0.97	0.30	0.72	0.75	0.77	0.80	0.83	0.85	0.88	0.91	0.94
	21	Pareto	(0.7,10)	0.97	0.10	0.90	0.91	0.92	0.92	0.93	0.94	0.95	0.95	0.96

- **Here α and β are the shape and the scale parameters, respectively**
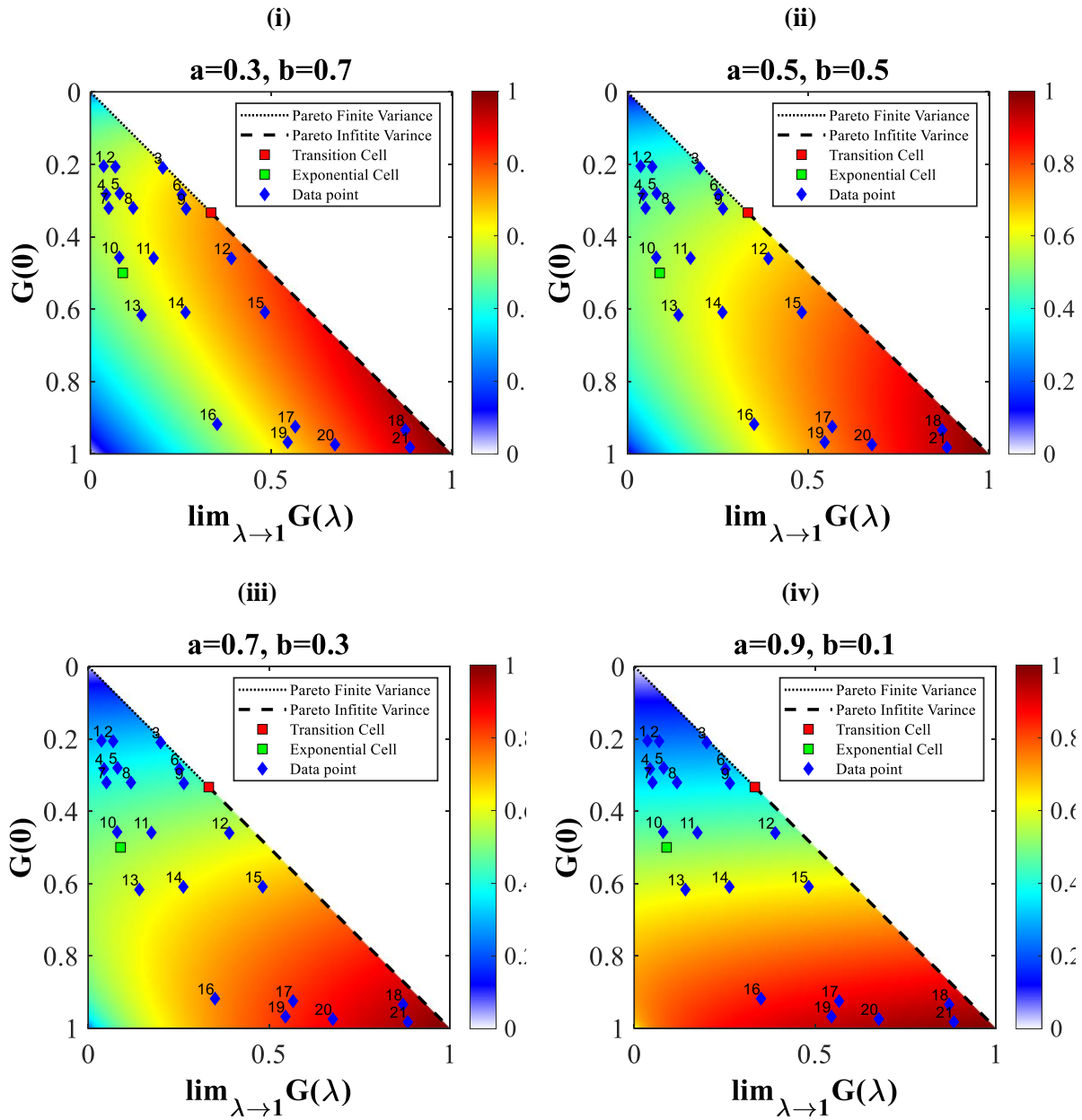
Figure 6.1 Concentration Maps with underlying Cobb-Douglas risk functions for different combinations of parameters a and b like (i) a=0.3, b=0.7, (ii) a=0.5, b=0.5, (iii) a=0.7, b=0.3 and (iv) a=0.9, b=0.1. The different points represent theoretical Concentration Profiles of useful distributions: Pareto, Lognormal, and Weibull. (Refer to Table 1 for information related to the type of distribution and its parameter represented in these CMs)

6.4   Study area and Data used

6.4.1   Observed Rainfall Data

The daily gridded observed rainfall data is obtained from the India Meteorological Department (IMD) for the period 1901-2020 (Pai et al., 2014). This data set of 0.25° latitude × 0.25° longitude resolution was developed using observed records of daily rainfall data from 6,995 rain-gauge stations across India. Vinnarasi and Dhanya, 2016; Mishra, et al., 2020; Gupta and Chavan, 2021; Shah and

Sateesh, 2022; Konda and Vissa, 2022. The entire range of data is split into two parts, viz. Pre-1970 period (1901-1970) and Post-1970 (1971-2018) to capture the changes (if any) in precipitation extremes. The year 1970 is chosen as the transition year in several earlier studies which detected some abrupt, substantial, and persistent changes or "shifts" in the state of natural climatic systems around the mid-to late-1970s in different parts of the globe, as well as in India (Miller et al., 1994; Baines and Folland, 2007; Sabeerali et al., 2012; Aadhar and Mishra, 2020; Sarkar and Maity, 2021; Gupta and Chavan, 2021, Gupta and Chavan, 2022). This phenomenon is generally referred to as the global shift in climate regime in the 1970s in literature, and its impact is manifested in several hydroclimatic variables, such as temperature, air pressure, wind field, precipitation, etc. Hence, a comparative analysis between the pre-and post-1970 period by choosing the year 1970 as the change year is expected to capture the potential impact of this 'global shift in climate regime in 1970s' on precipitation extremes over India.

6.4.2   *Simulated Future Precipitation Data*

For future analysis, this study considered simulated daily precipitation datasets (2020-2100) from four Shared Socio-economic Pathways (SSPs) - SSP126, SSP245, SSP370, and SSP585 - obtained from 16 GCMs participating in the Coupled Model Intercomparison Project-6 (CMIP6) (URL: https://esgf-node.llnl.gov/search/cmip6/accessed in January 2023) (see Table 6.2) (O'Neill et al., 2016; Eyring et al., 2016). The four distinct SSP scenarios of CMIP6 represent various levels of global development and greenhouse gas emissions. The SSP126 scenario envisions a sustainable world with strong climate policies and low population growth, while SSP245 portrays a moderate scenario with medium population growth and average climate policies (also termed a "middle of the road" scenario). SSP370 describes a fragmented world with high inequality and low economic growth, whereas SSP585 represents a world with high economic growth, high energy demand, rapid technological advances, and high greenhouse gas emissions. Each of these scenarios incorporates a variety of assumptions about future socio-economic conditions, such as population growth, scientific and technological advancements, consumptive energy uses, and land-use change. The SSP scenarios serve as a framework for understanding a range of possible climate futures based on different socio-economic pathways. Herein, we have selected 16 GCMs available under r1i1p1f1 initial conditions and five scenarios: historical, SSP126, SSP245, SSP370, and SSP585. GCM Models chosen in this study are found to display promising capabilities in making reliable assessments of the impacts of climate change on precipitation in Indian regions (Aadhar and Mishra, 2020; Gupta et al., 2020; Rajbanshi and Das, 2021; Chowdhuri et al., 2021; Prajeesh et al., 2022; Saha and Sateesh, 2022; Bhattacharya et al., 2022; Deepthi and Sivakumar, 2022; Vinod and Agilan, 2022; Konda and Vissa, 2023; Reddy and Saravanan, 2023).

The GCM datasets have considerable biases present in them due to model structural uncertainties, incomplete understanding of climate processes, uncertainties in external forcings, data assimilation processes, and spatial and temporal resolution (Mishra et al., 2014; Ghosh and Mujumdar, 2009; Gupta and Chavan, 2022). Simulated precipitation datasets are needed at suitable resolution for hydrological

analysis. Several methods are available to bias correct and downscale the climate variables like linear scaling, Quantile-Quantile mapping, Empirical Quantile Mapping, Semi-Parametric quantile mapping, Change factor method, Hybrid delta method and copula-based RMPH technique (Piani et al., 2010; Anandhi et al., 2011; Choudhary and Dimri, 2019; Mishra et al., 2020; Suman et al., 2021; Jaiswal et al., 2022; Oruc, 2022; Vishnupriya and Agilan, 2022; Rajulapati and Papalexiou, 2023). In this study, we considered the multiplicative multiple change factor method, a standard downscaling technique for simulating precipitation to match the resolution of the observed dataset (i.e., $0.25° \times 0.25°$), for facilitating the multimodel analysis. The historical simulations are obtained for the base-period (1984-2014) and future scenarios for all SSPs are obtained with respect to this. The SSP scenarios serve as a framework for understanding a range of possible climate futures based on different socio-economic pathways. Thus, by comparing changes between these different scenarios the possible impact of higher anthropogenic activity and greenhouse gas emissions can be assessed.

6.4.3   *Meteorological Subdivisions of India*

The climatic variables tend to vary considerably over India due to its vast spatial extents. To understand the regional climatic features, the study at the meteorological subdivision level might be useful. Hence, the meteorological subdivisions are generally used for research and monitoring purposes. The India Meteorological Department (IMD) has divided the country into 36 meteorological subdivisions (or coherent regions) for various applications (Guhathakurta and Rajeevan, 2008). These subdivisions are primarily meant for generating weather forecasts for political regions like states. Out of 36, 34 conterminous subdivisions were considered in the present study (Gupta and Chavan, 2022). (See Figure 3.1 in Chapter 3)

6.4.4   *Köppen-Geiger climate classification*

The Köppen-Geiger (KG) climate classification is a widely used technique that divides the climate across the globe into multiple categories based on temperature and precipitation patterns. The five major categories of climate types include A (Tropical), B (Dry), C (Temperate), D (Snow), and E (Polar), which can further be categorized into different subtypes based on precipitation and temperature patterns within each group, denoted by a small letter at the second and third place, respectively. India has a diverse range of climatic subtypes according to the Köppen system (see Figure 6.2). The Köppen-Geiger climate classification maps used in this study were prepared using data from the Climatic Research Unit (CRU) at the University of East Anglia and the Global Precipitation Climatology Centre (GPCC) at the German Weather Service and consider variables such as mean precipitation, temperature, and vegetation (Rubel and Kottek, 2010). India covers four of the five main climate classes: A, B, C, and E.

Table 6.2 Basic details of CMIP6 GCMs used in this study

S. No.	Model Name	Source Institute	Actual resolution (latitude × longitude)	Resolution after regridding (latitude × longitude)
1	ACCESS-CM2	Commonwealth Scientific and Industrial Research Organization, Australia	$1.25° \times 1.875°$	$0.25° \times 0.25°$
2	ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Organization, Australia	$1.25° \times 1.875°$	$0.25° \times 0.25°$
3	BCC-CSM2-MR	Beijing Climate Center, China	$1.11° \times 1.125°$	$0.25° \times 0.25°$
4	CESM2-WACCM	National Center for Atmospheric Research, USA	$0.9424° \times 1.25°$	$0.25° \times 0.25°$
5	CMCC-CM2-SR5	Fondazione Centro Euro-Mediterraneo sui CambiamentiClimatici, Italy	$0.9424° \times 1.25°$	$0.25° \times 0.25°$
6	CMCC-ESM2	Fondazione Centro Euro-Mediterraneo sui CambiamentiClimatici, Italy	$1.3° \times 1°$	$0.25° \times 0.25°$
7	EC-Earth3	European Centre for Medium-Range Weather Forecasts	$0.70° \times 0.70°$	$0.25° \times 0.25°$
8	EC-Earth3-Veg	European Centre for Medium-Range Weather Forecasts	$0.70° \times 0.70°$	$0.25° \times 0.25°$
9	GFDL-ESM4	NOAA Geophysical Fluid Dynamics Laboratory, USA	$1° \times 1.3°$	$0.25° \times 0.25°$
10	INM-CM4-8	Institute for Numerical Mathematics, Russian Academy of Science, Russia	$1.5° \times 2.0°$	$0.25° \times 0.25°$
11	INM-CM5-0	Institute for Numerical Mathematics, Russian Academy of Science, Russia	$1.5° \times 2.0°$	$0.25° \times 0.25°$
12	IITM	Indian Institute of Tropical Meteorology Pune, India	$1.9048° \times 1.8750°$	$0.25° \times 0.25°$
13	IPSL-CM6A-LR	Institut Pierre Simon Laplace, France	$1.2676° \times 2.5°$	$0.25° \times 0.25°$
14	MIROC6	JAMSTEC, AORI, NIES, and R-CCS, Japan	$1.41° \times 1.41°$	$0.25° \times 0.25°$
15	MR1-ESM2-0	Meteorological Research Institute, Japan	$1.11° \times {\sim}1.125°$	$0.25° \times 0.25°$
16	TaiESM1	Research Center for Environmental Changes, Academia Sinica, Taiwan	$1° \times 1.3°$	$0.25° \times 0.25°$

Detailed information about the type of conditions prevailing within each climate zone and the meteorological regions included within each climate zone is mentioned in Table 6.3. Of the major climate groups, a large proportion (33.03%) of stations were assigned to major climate type A, followed by climate type C (29.88%), B (27.76%), and least by climate type E (9.3%). Nine sub-climate zones, as defined using the Köppen-Geiger method, are used to subdivide the country (refer to Figure 6.2). The use of climate zones along with that of the meteorological region is done while conforming to the fact that weather and climate do not respect political boundaries, and climate zones might present a better understanding of the changes happening in the country (Gunwani and Mohan, 2017; Yaduvanshi et al., 2021).
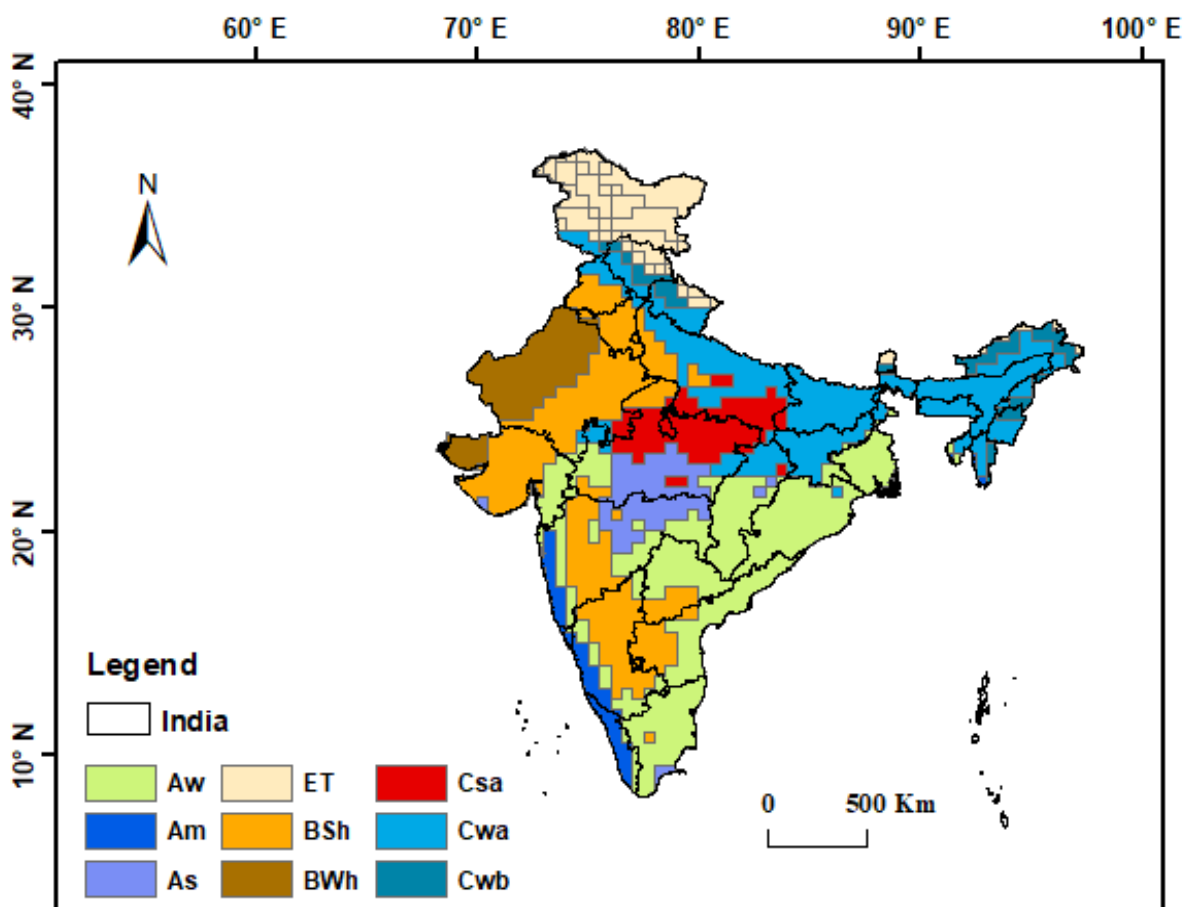
Figure 6.2 Maps showing the Köppen-Geiger climate zones of India (Rubel and Kottek, 2010)

Table 6.3 Detail related to the climate zones prevailing over India based on the Köppen-Geiger climate classification

Climate Type	Climate Zones	Regions
Tropical	Aw	Tropical wet savanna dry winter region includes Southern and Southern Eastern states of India, Orissa, Chattisgarh, Jharkhand, Tamil Nadu, and some parts of Andhra Pradesh, Telangana
	Am	Tropical monsoonal region, including Western Coastal Area
	As	Tropical dry savanna includes Parts of Madhya Pradesh and Maharashtra.
Arid	BSh	Semi-Arid regions with hot summers include Parts of Rajasthan, Gujarat, Punjab, Haryana, Delhi, Maharashtra, Karnataka, and other south Indian states
	BWh	Arid regions with extremely hot summers include Western part of Rajasthan and Saurashtra, Kutch and Diu
Temperate	Cwa	Temperate dry winter and hot summer includes NorthEastern and Eastern part of India
	Cwb	Temperate dry winter and warm summer includes Eastern Part of North East Indian states and some part of Himachal and Uttarakhand
	Csa	Temperate dry summer and hot summer includes Part of Uttar Pradesh and Madhya Pradesh
Polar	ET	Polar tundra snowfall in winter with cool summer includes the Mountainous region of Jammu& Kashmir, Ladhak, Himachal Pradesh, and Uttarakhand

6.5   *Results and Discussion*

6.5.1   *Observational changes in the tail risk of the daily gridded precipitation*

Using the approach of CM, this study evaluated and analyzed the risk dispersed in the tail of the probability distribution of the gridded precipitation dataset over India for two observational periods i.e., pre-1970s (1901-1969) and post-1970 period, also assessed the changes happening over time. Following the methodology discussed in section 6.2.1, initially, a CP is obtained for nearly 4801 gridded datasets and relevant information related to the risk drivers (r_1 and r_2) is extracted from each profile. Following the methodology, a concise risk score is calculated for each grid based on the Cobb-Douglas risk function for parameter values a=0.3 and b=0.7 for pre as well as post-1970s. Figure 6.3 shows the Concentration Map based on the Cobb-Douglas risk function for the daily precipitation dataset for the pre-post 1970s. Risk function values usually lie between 0 and 1. Higher the risk function value, the riskier the CP. Since the approach of CM is a scale-free, non-dimensional approach, it becomes easy to study the spatial variation of the risk associated with datasets and compare them over geographically diverse environments like the Indian mainland.

(i) Pre-1970s   **(ii) Post-1970s**
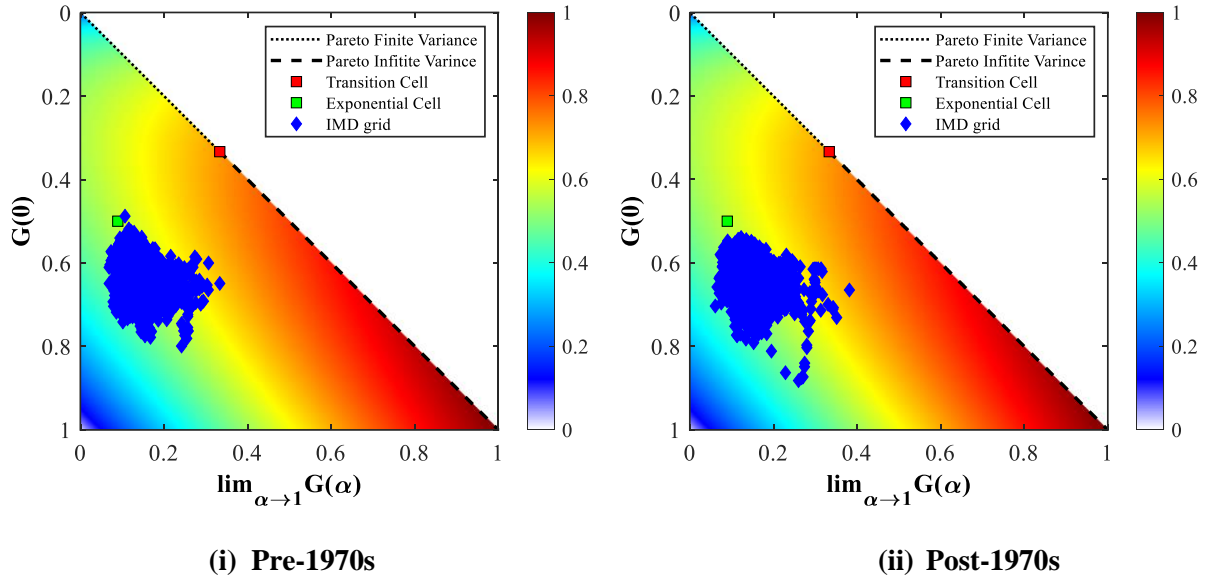
Figure 6.3 Risk evaluation of the daily precipitation datasets sets over India. Concentration Maps are shown for (i) Pre-1970s, and (ii) Post-1970s time periods. Cobb-Douglas parameters considered are a=0.3, b=0.7

Figure 6.4 shows the spatial variation of the tail risk across the Indian mainland for daily gridded precipitation data for the pre-post-1970s period (observational). The figure also depicts the spatial variation of changes in tail risk after the transition in 1970, along with the nature of the risk type (i.e., increasing or decreasing) based on the changes (see Figure 6.4 (iii), (iv)). The estimates of risk function (R) for all grids range from 0.40 to 0.75 for both time periods. In the Pre-1970s period, higher values of tail risk were generally observed in the northern states like Jammu and Kashmir, Himachal, Uttrakhand, and Rajasthan along with states like Orissa, Jharkhand, West Bengal, Bihar, and Andhra Pradesh lying in the south-eastern part of the country. For the post-1970 period, nearly 57.57% (i.e., 2764 out of 4801) grids over the Indian mainland showed an increase in the risk function values, indicating a rise in the tail risk, which corresponds to an increase in the frequency of occurrences of precipitation extremes. Barring some pockets in north India (particularly in the Ladhak region, Himachal, and Punjab), the north-western region (Rajasthan, Gujarat), and the north-eastern region of India (Assam, Meghalaya, Tripura), the tail risk is found to increase in most of the places across the Indian mainland, especially in the Peninsular region, south-eastern coastal area (e.g., Andhra Pradesh, Orissa) and Central part of India (e.g., parts of Madhya Pradesh). Goswami et al. (2006) also noticed an increase in the magnitude and frequency of extreme rainfall events significantly over central India during the period 1951–2000.
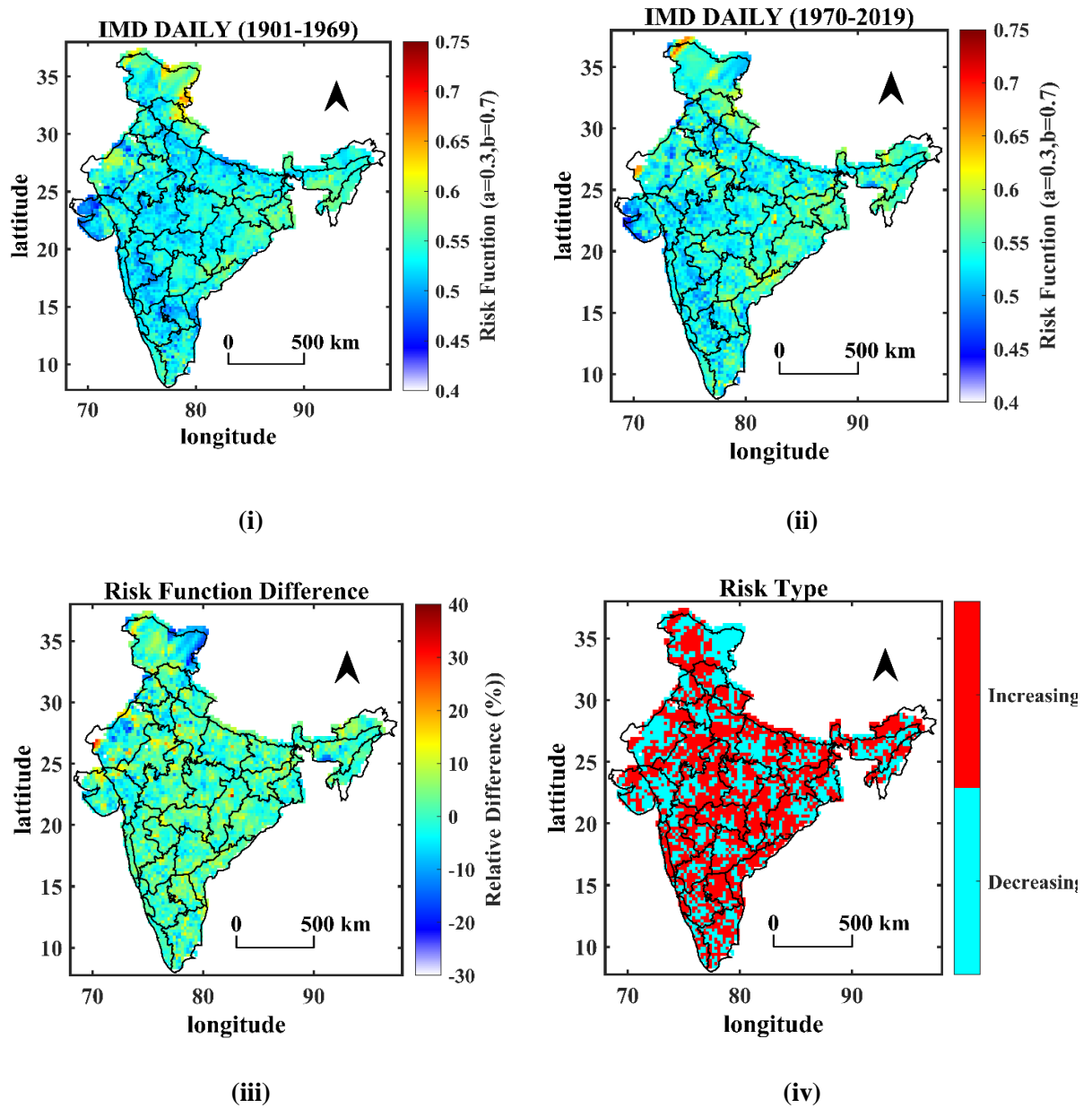
Figure 6.4 Spatial variation of (i) risk value for pre-1970, (ii) risk value for post-1970s, (iii) relative changes in risk in post-1970s wrt pre-1970s (%), and (iv) nature of the risk type based on the changes over time

Also, Dash and Maity (2019) noticed increasing trends in indices like maximum 1-day precipitation (RX1day), very wet day (R95p), and extremely wet day (R99p) precipitation in the southern part of India (peninsular, eastern, and west central), whereas patches of decreasing trends are located in northern India for the duration of post-1970s (i.e., 1976-2010). Sarkar and Maity (2020) also showed substantially higher estimates of Probable Maximum Precipitation (PMP) (that lies in the tail part) post-1970 compared to pre-1970 in the area similar to the one mentioned in this study, as a possible consequence of climate change and global climatic shift in the 1970s. Past findings are consistent with the pattern of tail risk noticed during the post-1970 in this study.

On the whole, the findings exhibit a notable dispersion in terms of change (i.e., increase or decrease); Yet overall a general pattern of increasing tail risk is observed across the Indian mainland during the post-1970 period. To further validate this finding over the entire Indian landmass, a Kolmogorov–Smirnov (KS) test (Massey, 1951; Marsaglia et al., 2003; Gupta and Chavan, 2022) was used to compare the distribution of tail risk during the pre-1970s period with its distribution during the post-1970s period (over nearly 4801 grid) at a 5% significance level. Figure 6.5 shows the cumulative distribution functions (CDFs) of risk function values for the two time periods i.e., pre-1970 (1901-1969) and post-1970 (1970-2019). The KS statistic value of 0.1056 (much higher than the critical value) suggests a considerable change in the distribution of the tail risk in the post-1970s with respect to the pre-1970s period.
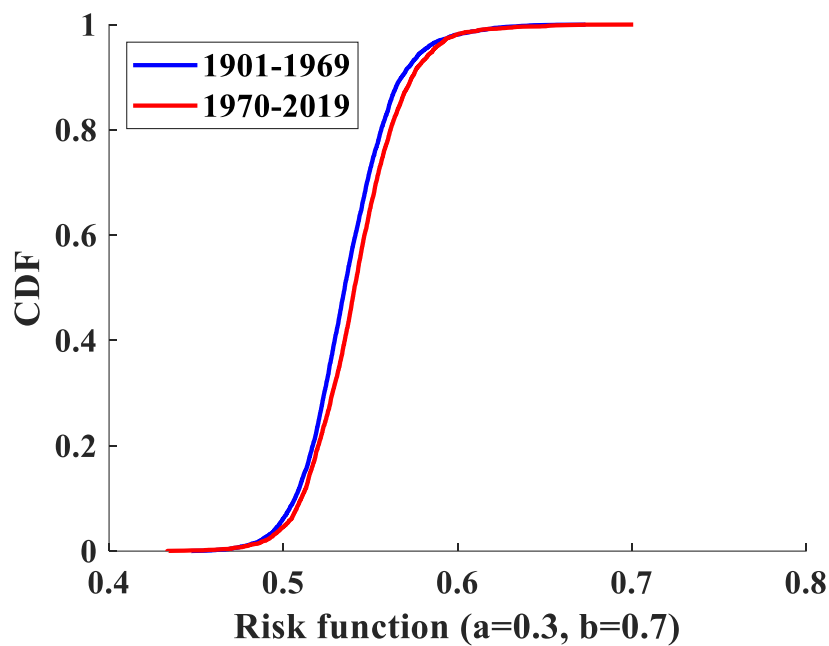
Figure 6.5 Plots showing the CDF of the risk function values over the Indian mainland estimated from Cobb-Douglas risk function corresponding to pre and post-1970s time period

Considerable changes in the tail risk show that the frequency and magnitudes of the extremes have changed over India in the post-1970s period which can be very well attributed to climate shifts that happened around the mid-to-late 1970s (Sarkar and Maity, 2020).

6.5.2 *Future-Projected changes in tail risk of the daily gridded precipitation: Multi-Model Assessment*

Risk function values based on the Cobb-Douglas risk function are evaluated for the downscaled future projections of the daily gridded precipitation datasets from 2020 to 2100 for all 16 CMIP6-GCMs. Further, the estimates of risk function were averaged at each grid to form a multimodel ensemble mean. Ensemble of the models was preferred over an individual model as it ensures a robust and more

homogeneous pattern of changes in climate state by for fending errors and internal variability of a single model (Yaduvanshi et al., 2021; Sarkar and Maity, 2022; Gupta and Chavan, 2022). These multimodel ensemble models have the ability to depict a more realistic pattern of future tail risk and can yield a more consistent pattern of changes happening over time in the future due to climate change across the Indian mainland. The results of the spatial distribution of the ensembled tail risk are shown in Figures 6.6, 6.7, 6.8, and 6.9 for four scenarios namely SSP126, SSP245, SSP370, and SSP585, respectively. The changes in the future tail risk are assessed with respect to the historical period of post-1970s (1970–2019). Visual interpretation of all the figures reveals that the spatial distribution of the multimodel ensemble mean of tail risk values in the future differs notably from the historical time period for all scenarios. Table 6.4 presents the summary of the number of grids over India showing either an increase or decrease in Relative risk (in %) for various climate change scenarios.

Table 6.4 Summary of Relative risk difference ranges and anticipated grid counts for various ranges. (The relative risk difference ranges are in Percentages)

		Relative risk difference					
Changes in future		**Decrease**			**Increase**		
Climate Scenarios	**-30 to -10**	**-10 to 0**	**Total grids**	**0 to 10**	**10 to 30**	**Total grid**	
SSP126	5	2060	2065	2736	0	2736	
SSP245	2	843	845	3956	0	3956	
SSP370	1	514	515	4263	23	4286	
SSP585	1	244	245	4338	218	4556	

It can be seen that for SSP126 (see figure 6.6), barring a few grids in the western Deccan plateau, the northern plain, and the eastern coastal plain almost all other grids i.e., nearly 56.99% (2736 grids out of 4801) showed an increase in the risk value in the future. The overall change presented in the form of relative risk difference (%) ranges from -10% decrease to a 5% increase in the future. For scenario, SSP245, the spatial variation of tail risk suggests few staggered grids with decreasing tail risk along the western coastal areas of Karnataka, Kerala, and in the north-eastern states of Tripura, Nagaland, Meghalaya, and Assam. The number of grids showing an increase in future rise to nearly 82.40% (3956 grids out of 4801) as compared to the historical period for this scenario (see Figure 6.7). Figure 6.8 shows the variation of the tail risk under the scenario of SSP370, where an increase in the number of grids showing higher tail risk is evident, especially in the Northern, Northwestern, Central, and Southeastern parts (eastern coast) of the country. Nearly 89.27% (4286 out of 4801) grids showed an increase in tail risk for the future time period, out of which nearly 740 grids showed a change of more than 5% or up to 12% in tail risk values. Further, we noticed that the most pronounced variations are observed undoubtedly in the SSP585 which is the worst-case scenario in terms of changing climate. Nearly 94.90% (4556 out of 4801) of grids all over the country showed an increase in the risk value. A

relative change in the tail risk value of up to 18% can be seen especially in the northern, western, and central parts of the country. The finding matches well with that of Sarkar and Maity (2022), who also noticed an overall increasing pattern of precipitation extremes based on the CMIP6 model over India. They observed a two-fold and three-fold increase in the frequency of 99th percentile daily rainfall extremes (i.e., tail part) for SSP245 and SSP585 than the historical period (1951-2020), respectively. For future time period, the western and northern parts of India show a higher extent of increase in the tail risk than the eastern part of the country. Sarkar and Maity (2020) also noticed substantial evidence regarding changes in PMP estimates within west-central India and parts of north-west India, where they found an overall increase in PMP estimates, as well as an overall decrease in PMP in eastern and southern parts of India for the future time period with respect to post-1970s. These changes in tail behavior over different regions of India can be attributed to the overall increased air temperature throughout India, which can augment the possibility of an increased moisture-holding capacity of the atmospheric column following the Clausius-Clapeyron relationship (Trenberth 2011), resulting in more extreme precipitation events over the northern and western parts of India. Moreover, a recent study by Singh et al. (2022) has clearly identified the increased surface pressure over the Tibetan plateau as a key factor contributing to changes in the precipitation patterns over north-west India.
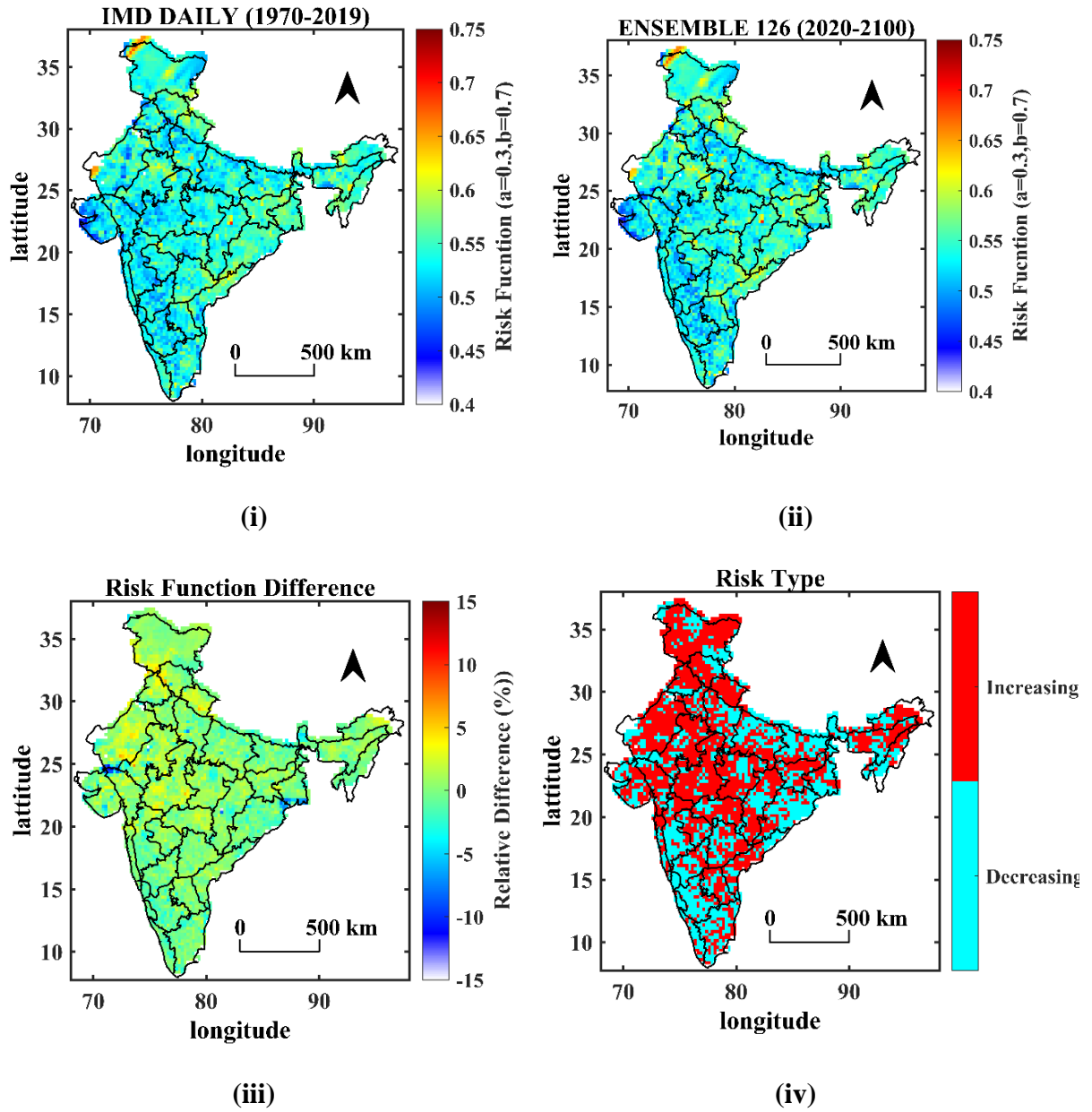
Figure 6.6 Spatial variation of (i) risk value for post-1970s, (ii) risk value from future-projected ensemble models for SSP126, (iii) percentage change (%) in future risk compared to post-1970s period, and (iv) characterization of risk type based on changes in future, across the Indian mainland
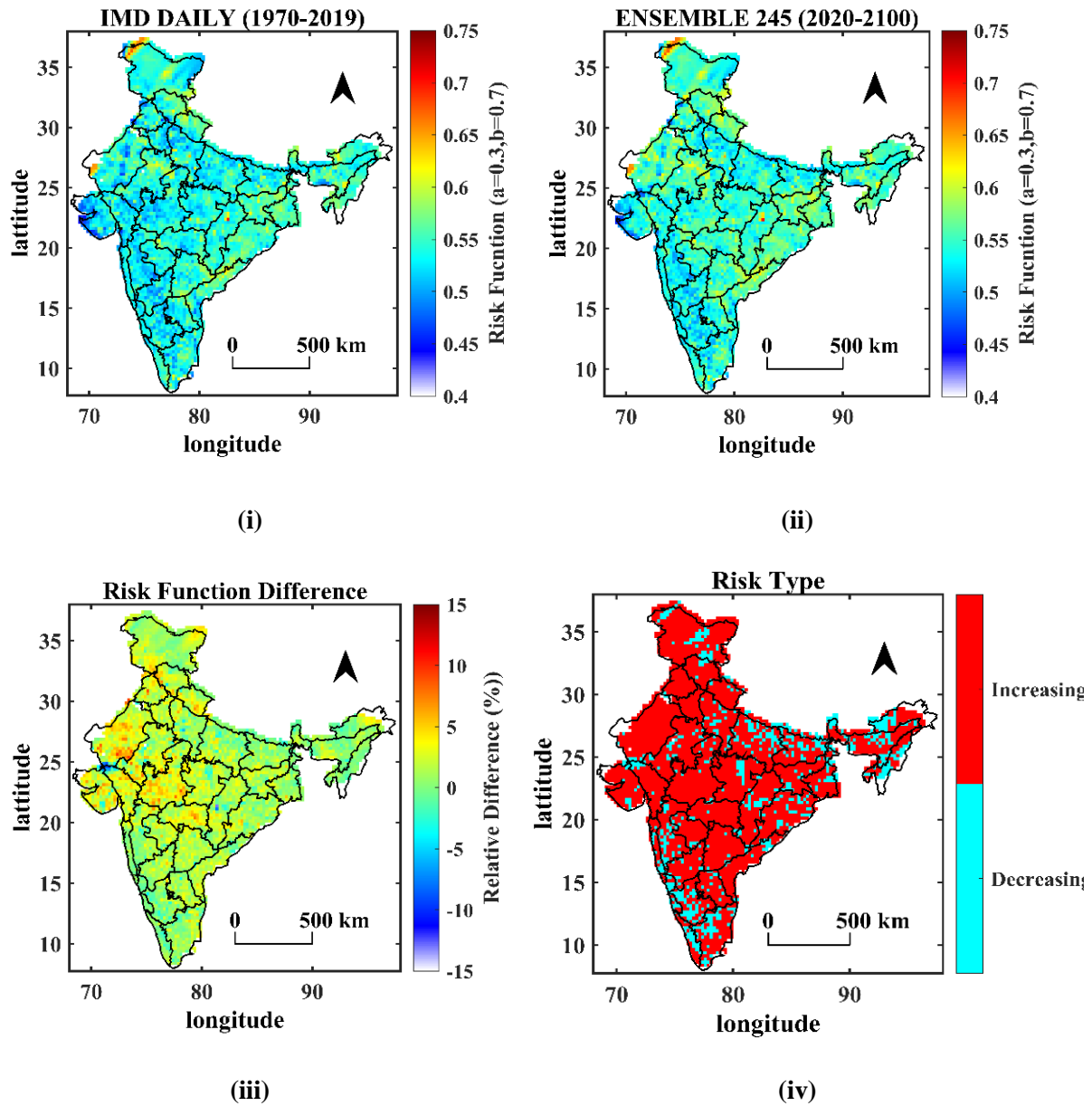
Figure 6.7 Spatial variation of (i) risk value for post-1970s, (ii) risk value from future-projected ensemble models for SSP245, (iii) percentage change (%) in future risk compared to post-1970s period, and (iv) nature of risk type based on changes in future, across the Indian mainland.
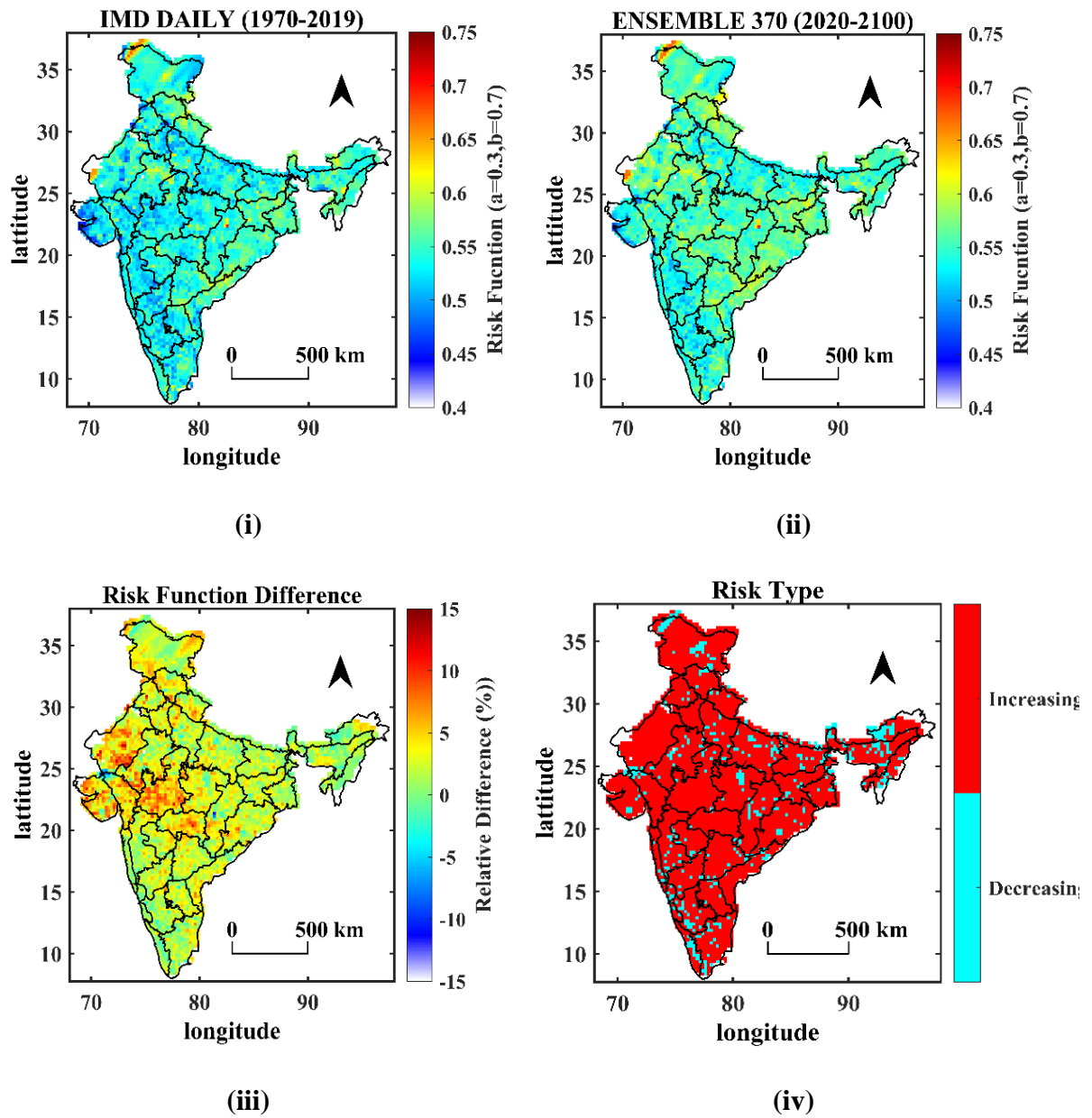
Figure 6.8 Spatial variation of (i) risk value for post-1970s, (ii) risk value from future-projected ensemble models for SSP370, (iii) percentage change (%) in future risk compared to post-1970s period, and (iv) nature of risk type based on changes in future, across the Indian mainland

IMD DAILY (1970-2019) (i)

ENSEMBLE 585 (2020-2100) (ii)

Risk Function Difference (iii)
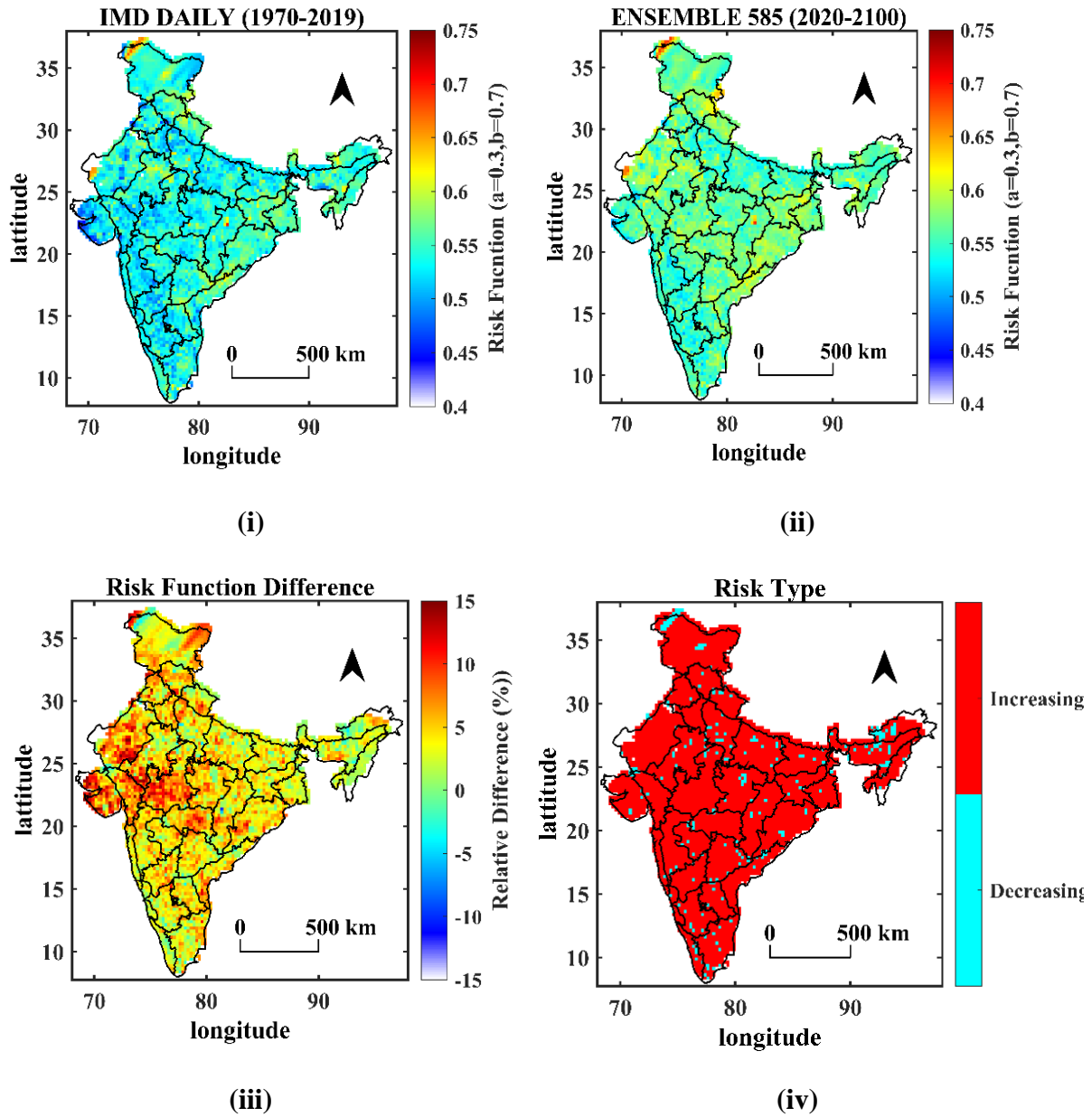
Risk Type (iv)

Figure 6.9 Spatial variation of (i) risk value for post-1970s, (ii) risk value from future-projected ensemble models for SSP585, (iii) percentage change (%) in future risk compared to post-1970s period, (iv) nature of risk type based on changes in future, across the Indian mainland

Further, to confirm the finding of increasing tail risk and assess whether the changes are significant or not for different scenarios in the future with respect to the historical period of post-1970, we compared the distribution of the tail risk of the historical time with that of the distribution corresponding to the 4 SSP scenarios (SSP126, SSP245, SSP370, and SSP585) using a KS test at 5% significance level. Figure 6.10 shows the plots of the CDFs of the ensemble of risk values formed from 16 CMIP6 GCM models for four climate scenarios and the CDF of risk values for historical time series (i.e., post-1970s). The KS statistic for scenarios SSP126, SSP245, SSP370, and SSP585 takes the estimates as 0.0285, 0.1445, 0.2580, and 0.4222, respectively, which are found to be significant at a 5%

154

significance level. Overall from the visual interpretation of the spatial patterns of tail risk (Figure 6.6 – 6.9), as well as from the CDF plots and KS statistics, it can be noted that the most pronounced variations are observed in the SSP585 (worst-case scenario) while the least significant changes can be seen in the SSP126 (sustainable scenario) at grid level over entire India. This observation is in line with the findings of other studies that also confirm that the role of anthropogenic forcing is more prominent for increased extremes in the future over India (Mukherjee et al., 2018; Reddy and Saravanan, 2023). It is to be noted that the absence of a study related to tail risk for the historical as well as future climate change scenarios over India restricts our ability to draw comparisons in our study; however, reasonably similar results from other studies focusing on the frequency of occurrences of extreme events (e.g., 95th or 99th percentiles) on India also supports the findings regarding the tail risk over India.

Figure 6.10 Plots showing the comparison of the CDFs of risk values obtained from ensembled CMIP6 GCMs corresponding to four climate scenarios, namely SSP126, SSP245, SSP370, and SSP585 with the CDF of risk values obtained for historical time period i.e., post-1970s over Indian mainland

6.5.3 Tail Risk assessment of gridded precipitation over the Meteorological Subdivision of India

The application of the CM approach helps in evaluating the tail risk associated with the probability distribution of the gridded precipitation dataset over India for both observational as well as future time period. However, understanding the spatial and temporal shifts can be challenging due to the variable patterns emerging in different regions in different time periods. Consequently, to enhance the applicability of the information gained at the gridded level for local governance purposes — such as facilitating reliable planning for the design and maintenance of hydrological structures — we have undertaken an analysis at the meteorological subdivisions level.

6.5.3.1 *Observational Changes in Meteorological Subdivisions*

The analysis at the subdivision level is performed by extracting the tail risk values at the grids falling within the specified meteorological region for the observational time period i.e., pre-1970s and post-1970s, and then comparing their distribution for changes over time. KS test was used to test the null hypothesis that the distribution of the tail risk for a meteorological subdivision has not changed from the pre-1970s to the post-1970s. For brevity, the CDF plots of risk values for two meteorological subdivisions namely Arunachal Pradesh and Kerela for pre and post-1970s are presented in Figure 6.11. Details related to the KS statistic along with the result from the KS test are also presented in Table 6.5.
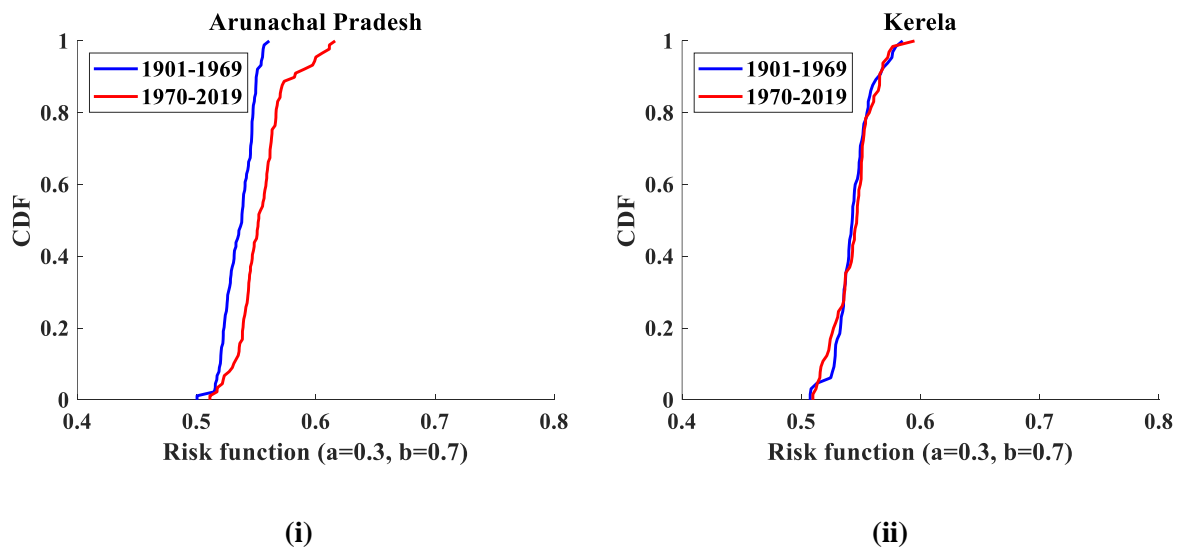
Figure 6.11 Plots showing the CDFs of the risk values over two meteorological subdivisions namely (i) Arunachal Pradesh and (ii) Kerela, for two observational time periods i.e., pre-1970s (1901-1969) and post-1970 (1970-2019)

Table 6.5 Statistics of KS test and the result from hypothesis testing based on KS test for meteorological subdivisions assessing the changes happening between the pre and post-1970s period. (Here, h=1 means to reject the null hypothesis, and h=0 means to accept the null hypothesis).

Subdivision	KS test	
	KS statistic	h
Arunachal Pradesh	0.4607	1
Assam and Meghalaya	0.1018	0
Naga Mani Mizo and Tripura	0.1222	0
Sub Him W Bengal Sikkim	0.1935	0
Gangetic West Bengal	0.2128	1
Orissa	0.1682	1
Jharkhand	0.3158	1
Bihar	0.1346	0
East Uttar Pradesh	0.1773	1
West Uttar Pradesh	0.2338	1
Uttaranchal	0.2530	1
Haryana Chandigarh and Delhi	0.2286	1
Punjab	0.1685	0
Himachal Pradesh	0.2386	1
Jammu and kashmir	0.2642	1
West Rajasthan	0.0903	0
East Rajasthan	0.2573	1
West Madhya Pradesh	0.2308	1
East Madhya Pradesh	0.0765	0
Gujarat	0.1774	1
Saurashtra Kutch and Diu	0.1656	1
Konkan and Goa	0.3143	1
Madhya Maharastra	0.1908	1
Marathwada	0.2584	1
Vidarbha	0.4453	1
Chhatisgarh	0.2333	1
Coastal Andhra Pradesh	0.2560	1
Telangana	0.2000	1
Rayalaseema	0.3778	1
Tamil Nadu and Pondicherry	0.1086	0
Coastal Karnataka	0.3077	1
North Interior Karnataka	0.2404	1
South Interior Karnataka	0.1301	0
Kerela	0.1385	0

Based on the test results, we found that nearly 22 subdivisions out of 34 showed significant changes in tail risk between the pre- and post-1970s. Subdivisions like Arunachal Pradesh, Orissa, Jharkhand, Eastern Uttar Pradesh, Western Uttar Pradesh, Uttaranchal, Himachal Pradesh, East Rajasthan, West Madhya Pradesh, Konkan and Goa, Madhya Maharashtra, Marathwada, Vidharbha, Chhattisgarh, Coastal Andhra Pradesh, Telangana, Rayalaseema, Coastal Karnataka, and North Interior Karnataka showed significant increasing changes in the post-1970s period relative to pre-1970s. While subdivisions like Gangetic West Bengal, Haryana Chandigarh and Delhi, Jammu and Kashmir, Gujarat, and Saurashtra Kutch and Diu also showed a change in the distribution of tail risk; however, the changes were of decreasing nature. The subdivisions that showed a tendency towards an increase in the tail risk can be prone to frequent occurrences of precipitation extremes.

6.5.3.2 *Future changes corresponding to four scenarios in Meteorological Subdivisions*

Similar to the analysis performed in the historical period, the meteorological subdivision-wise analysis is also performed for future time periods (2020-2100) considering the climate change scenarios to investigate how the frequency of precipitation extremes would be in the subdivisions. Again, a KS test at a 5% significance level was employed to verify the considerable change in the distribution of tail risk for a particular subdivision based on the estimates obtained from ensembled GCMs for estimates corresponding to the historical time period (herein post-1970s). Figure 6.12 shows the CDF plots of the risk estimates obtained from an ensemble of 16 CMIP6 GCMs, along with that of the historical time period (i.e., post-1970s) for all 34 subdivisions. Table 6.6 provides the KS statistics along with the result from the KS test for all meteorological subdivisions corresponding to SSP126, SSP245, SSP370, and SSP585 climate scenarios. Based on the CDF plots and KS statistics following inferences can be drawn. Under the climate scenario SSP126, only 3 out of 34 subdivisions showed significant changes in the distributions of tail risk. For climate scenario SSP245, 19 out of 34 subdivisions showed significant changes. Notably, the CDF for the future time period usually lies beyond the CDFs of the historical time periods of the 1970s indicating an increasing tendency of tail risk. The maximum increase in the tail risk is observed for the subdivision of Vidharbha which is followed by West Madhya Pradesh, Marathawada, Haryana, Chandigarh and Delhi, West Rajasthan, East Madhya Pradesh, Gujarat, East Rajasthan, Punjab, Himachal Pradesh, Chhattisgarh, Uttaranchal, Telangana, Arunachal Pradesh, Gangetic West Bengal, Jammu and Kashmir, Saurashtra Kutch and Diu, Coastal Andhra Pradesh, and Orissa. Turning attention to the SSP370 scenario, changes in the CDFs of tail risk for the future are observed in 30 out of 34 subdivisions. The maximum changes (increasing) are seen in the subdivision of West Madhya Pradesh followed by Vidharba, Gujarat, East Madhya Pradesh, Marathwada, Haryana Chandigarh and Delhi, Chhattisgarh, West Rajasthan, Saurashtra Kutch and Diu, Orissa, East Rajasthan, Jammu and Kashmir, Himachal Pradesh, Arunachal Pradesh, Coastal Andhra Pradesh, Rayalaseema, Jharkhand, Telangana, Gangetic West Bengal, Uttaranchal, West Uttar Pradesh, Punjab, East Uttar Pradesh, Konkan and Goa, Madhya Maharashtra, North Interior Karnataka, South Interior Karnataka, Bihar,

158

Tamil Nadu and Pondicherry and lastly Assam and Meghalaya. Further moving to the SSP585 scenario, all subdivision shows significant changes in the CDFs of the tail risk. Each subdivision indicated an increased tail risk particularly prominent in Gujarat followed by West Madhya Pradesh, Vidarbha, East Madhya Pradesh, Chhattisgarh, Orissa, Marathawada, Haryana Chandigarh and Delhi, Saurashtra, Kutch and Diu, Gangetic West Bengal, Konkan and Goa, Jammu and Kashmir, West Rajasthan, East Rajasthan, Jharkhand, Coastal Andhra Pradesh, North Interior Karnataka, Madhya Maharastra, Punjab, Himachal Pradesh, Rayalseema, Telangana, Coastal Karnataka, South Interior Karnataka, Bihar, West Uttar Pradesh, Arunachal Pradesh, East Uttar Pradesh, Uttaranchal, Tamil Nadu and Pondicheery, Kerela, Sub Him West Bengal Sikkim, Assam, and Meghalaya and Naga Mani Mizo and Tripura.
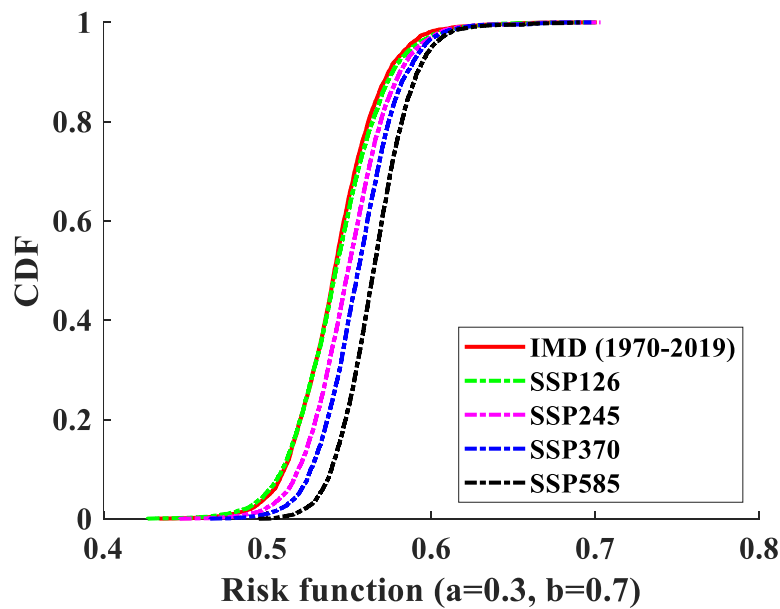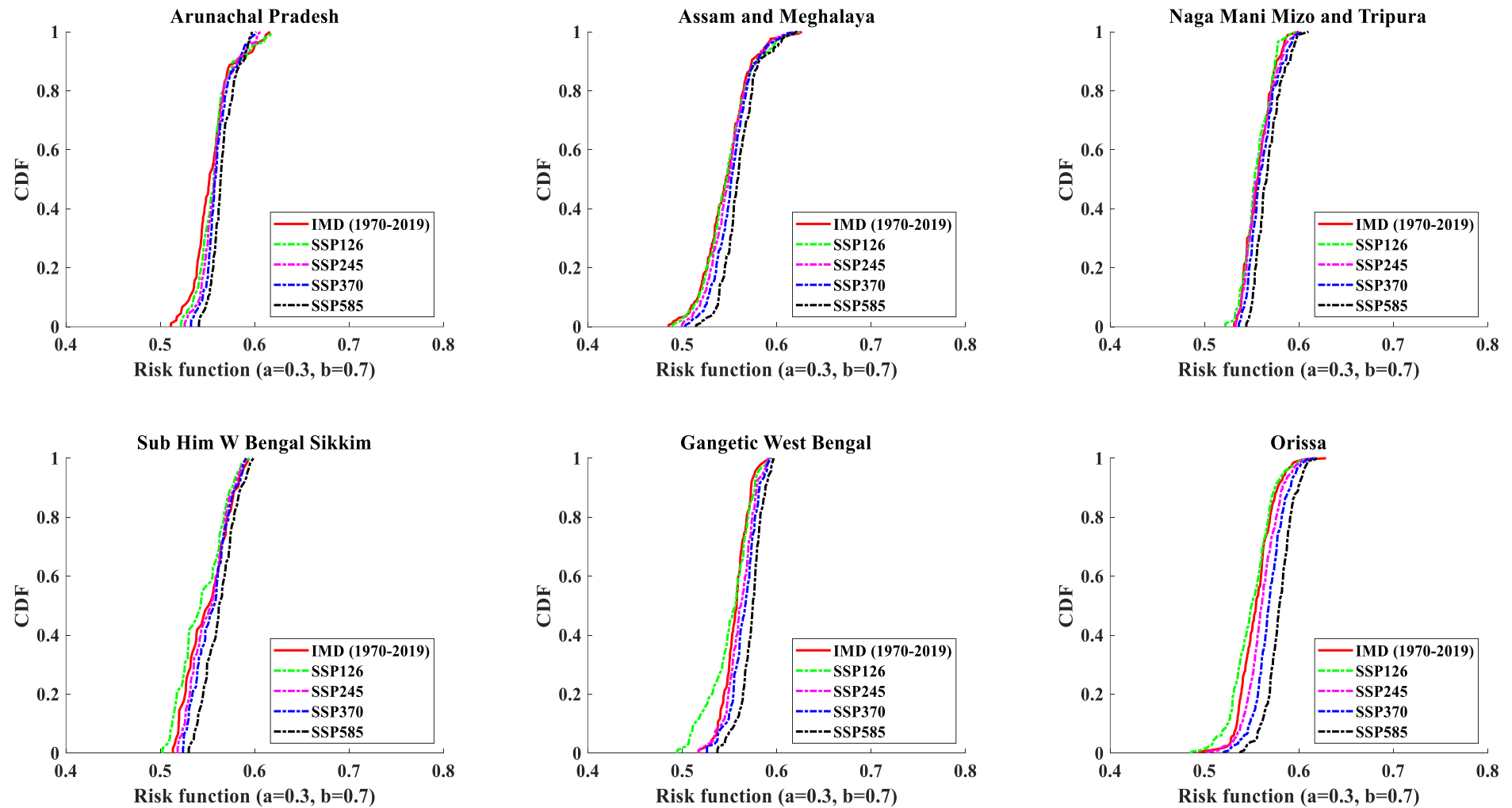
Figure 6.12 Comparison of the CDFs of the risk values obtained from the ensemble of 16 CMIP6 GCMs corresponding to four climate scenarios, namely SSP126, SSP245, SSP370, and SSP585 with the CDF obtained for post-1970s for 34 meteorological subdivisions

(**Figure 6.12** Continued…..)
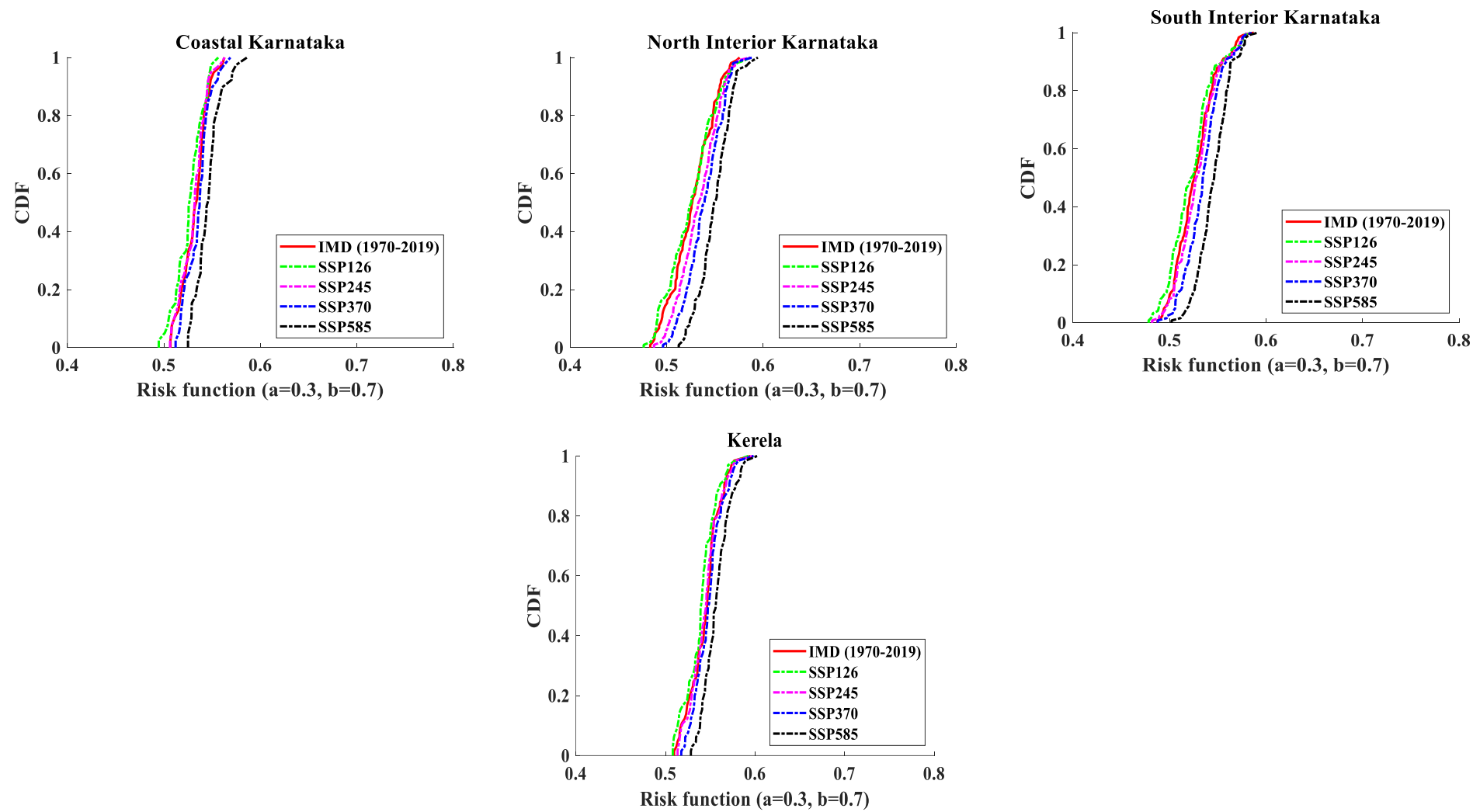
Table 6.6 Statistics of KS test and the result from hypothesis testing based on KS test for meteorological subdivisions assessing the changes happening in the future corresponding to different climate scenarios with respect to post-1970s period (Here, h=1 means to reject the null hypothesis and h=0 means to accept the null hypothesis)

Subdivisions	KS test							
	SSP126		SSP245		SSP370		SSP585	
	KS statistic	h	KS statistic	h	KS statistic	h	KS statistic	h
Arunachal Pradesh	0.1348	0	0.2022	1	0.3146	1	0.4157	1
Assam and Meghalaya	0.0479	0	0.0898	0	0.1737	1	0.3054	1
Naga Mani Mizo and Tripura	0.0889	0	0.0778	0	0.1889	0	0.3000	1
Sub Him W Bengal Sikkim	0.1613	0	0.1129	0	0.1774	0	0.3226	1
Gangetic West Bengal	0.1915	0	0.2021	1	0.2872	1	0.5532	1
Orissa	0.1542	1	0.1776	1	0.3551	1	0.5888	1
Jharkhand	0.0965	0	0.1667	0	0.2982	1	0.4737	1
Bihar	0.0962	0	0.0833	0	0.2051	1	0.4295	1
East Uttar Pradesh	0.0636	0	0.1182	0	0.2091	1	0.4091	1
West Uttar Pradesh	0.0519	0	0.1429	0	0.2662	1	0.4286	1
Uttaranchal	0.1566	0	0.2169	1	0.2771	1	0.3976	1
Haryana Chandigarh and Delhi	0.1143	0	0.3000	1	0.4000	1	0.5714	1
Punjab	0.1685	0	0.2360	1	0.2584	1	0.4607	1
Himachal Pradesh	0.1477	0	0.2273	1	0.3182	1	0.4545	1
Jammu and Kashmir	0.0938	0	0.1951	1	0.3333	1	0.5160	1
West Rajasthan	0.1155	1	0.2924	1	0.3935	1	0.5162	1
East Rajasthan	0.0874	0	0.2379	1	0.3350	1	0.5049	1
West Madhya Pradesh	0.1417	1	0.3077	1	0.5142	1	0.6883	1
East Madhya Pradesh	0.0969	0	0.2908	1	0.4592	1	0.6378	1
Gujarat	0.0645	0	0.2742	1	0.4839	1	0.7097	1
Saurashtra Kutch and Diu	0.0382	0	0.1911	1	0.3631	1	0.5669	1
Konkan and Goa	0.0714	0	0.1286	0	0.2429	1	0.5286	1
Madhya Maharastra	0.0855	0	0.1382	0	0.2368	1	0.4671	1
Marathwada	0.1124	0	0.3034	1	0.4157	1	0.5730	1
Vidarbha	0.1314	0	0.3431	1	0.5036	1	0.6496	1
Chhatisgarh	0.0556	0	0.2222	1	0.3944	1	0.6056	1
Coastal Andhra Pradesh	0.0560	0	0.1840	1	0.3120	1	0.4720	1
Telangana	0.0533	0	0.2067	1	0.2933	1	0.4400	1
Rayalaseema	0.0667	0	0.1778	0	0.3000	1	0.4556	1
Tamil Nadu and Pondicherry	0.0800	0	0.1086	0	0.2000	1	0.3714	1
Coastal Karnataka	0.2308	0	0.1026	0	0.1538	0	0.4359	1
North Interior Karnataka	0.1058	0	0.1538	0	0.2308	1	0.4712	1
South Interior Karnataka	0.1382	0	0.0813	0	0.2114	1	0.4309	1
Kerela	0.2308	0	0.0769	0	0.1231	0	0.3692	1

6.5.4 Tail Risk assessment of the tails of gridded precipitation over the Köppen-Geiger regions of India

While administrative boundaries are important for a country to take action necessary for protection against the risk of precipitation extremes, an understanding of tail risk is also needed for different climate zones as the climate does not respect political boundaries. Thus, an understanding of tail risk was evaluated for the Köppen-Geiger (KG) classification which were formed based on the climatology (i.e., variability in precipitation and temperature) of the country. Figure 6.13 illustrates the variations in the risk values across all stations associated with each KG climate subtype encompassing India's mainland for the historical (1970-2019) and future (2020-2100) time periods corresponding to four different climate scenarios. The box plots were constructed to represent the ensemble spread of tail risk corresponding to each climate scenario for all 9 climate zones relative to spread for the historical time period (i.e., post-1970s). The statistical significance of change based on the distributions of ensembles for their respective climate scenario at a 5% significance level is evaluated using the KS test. Table 6.7 provides the KS statistics along with the results from the KS test for different Köppen regions.

Historically, the maximum tail risk values were observed for climate type Cwb and ET, followed by Aw, Cwa, Csa, BWh, Am, As, and BSh, respectively, based on their median values. The grids within Cwb and ET regions have higher elevations (hilly or mountainous regions), and receive a good amount of rainfall (see Figure S9 in supplementary material of Gupta and Chavan, 2022). Higher elevations are usually associated with the presence of some of the heaviest tails of daily or hourly precipitation datasets in several parts of the world as shown by Papalexiou et al. (2018), Hobbi et al. (2021), Gupta and Chavan (2022). Heavy tails usually suggest the higher frequency of extreme precipitation events thereby reflecting higher tail risk.

For the future time period, for different climate zones, the largest to the smallest value of tail risk was observed in BWh, followed by Aw, As, BSh, ET, Csa, Cwa, Cwb, and Am for all four SSP scenarios. A clear difference in the spatial pattern of future tail risk for precipitation over India can be seen as compared to the historical pattern. For the observational time period, Cwb and ET zones occupying some parts of northern and north-eastern India exhibited higher tail risks. However, in the future, BWh, Aw, As, and BSh, occupying the western, central, and south-eastern part of the country exhibits higher tail risk. Suman and Maity (2020), through their analysis, revealed an intensified eastward moisture flux over the Indian Ocean and Arabian Sea region resulting in an overall increase in extreme precipitation events in the southern part of the country in the future. Further, we also noticed that the grids associated with climate type Aw have the maximum inter-quantile range, which suggests a large dispersion or variability of tail risk within this region.
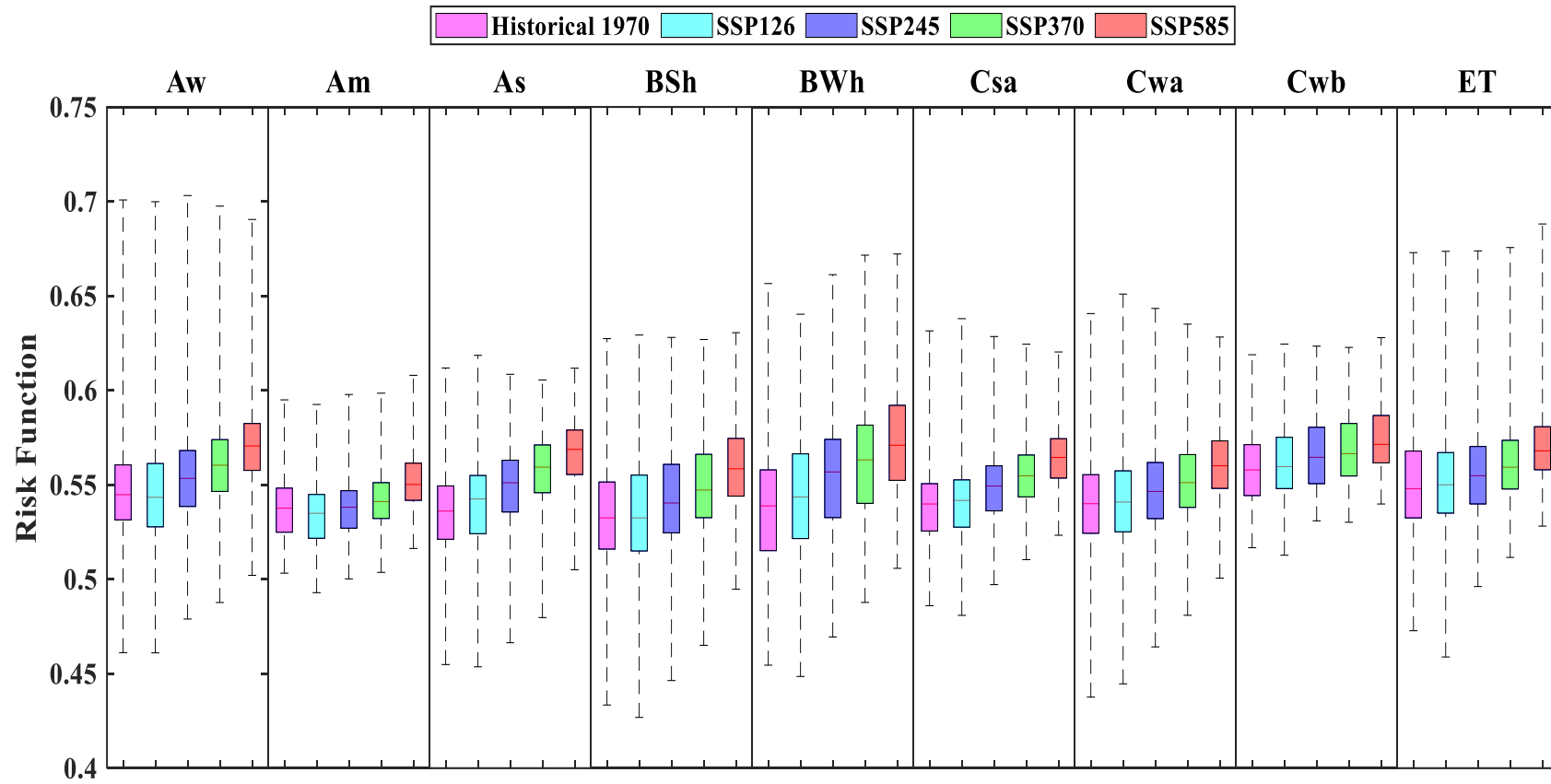
Figure 6.13 Variation of the tail risk value for daily precipitation time series associated with each climate type across India for both the historical (1970 to 2019) as well as future (2020-2100) period for different climate scenarios

Table 6.7 Statistics of KS test and the result from hypothesis testing based on KS test for Köppen-Geiger Regions assessing the changes happening in the future corresponding to different climate scenarios with respect to the post-1970s period (Here, h=1 means to reject the null hypothesis, and h=0 means to accept the null hypothesis)

Köppen-Geiger Regions	KS test							
	SSP126		SSP245		SSP370		SSP585	
	KS statistic	h	KS statistic	h	KS statistic	h	KS statistic	h
Aw	0.0613	1	0.1650	1	0.3021	1	0.4581	1
Am	0.1020	0	0.0561	0	0.1531	1	0.3673	1
As	0.1246	1	0.2918	1	0.4413	1	0.6192	1
BSh	0.0425	0	0.1454	1	0.2572	1	0.4273	1
BWh	0.0994	0	0.2391	1	0.3168	1	0.4503	1
Csa	0.0718	0	0.2348	1	0.3729	1	0.5773	1
Cwa	0.0332	0	0.1240	1	0.2292	1	0.3865	1
Cwb	0.0882	0	0.1765	1	0.2353	1	0.3412	1
ET	0.0515	0	0.1454	1	0.2617	1	0.4251	1

Further, Table 6.7 provides the KS statistics along with the results from the KS test for different Köppen regions. Results indicate statistically significant changes for climate zones As, and Aw for the SSP126 scenario. For scenario SSP245 all zones except Am showed significant changes from the historical time period and are ordered as As followed by BWh, Csa, Cwb, Aw, ET, BSh, and Cwa. For climate scenarios, SSP370 and SSP585 all climate zones showed significant changes in the future wrt historical time period. The climate zones for SSP370 are ordered such that As is followed by Csa, Bwh, Aw, ET, BSh, Cwb, Cwa, and Am in terms of changes from the post-1970s period. For, SSP585 the climate zones are ranked as As followed by Csa, Aw, BWh, BSh, ET, Cwa, Am, and Cwb from largest to smallest changes. It is to be noted that the biggest changes (increase) in the tail heaviness is seen primarily for the SSP585 scenario for most climate subtypes.

Based on the observed values of tail risk in the future and the changes in tail risk over time, we found that the Tropical (Aw, As), Arid zones (BWh, BSh), and Temperate (Csa) are found to be more susceptible to higher precipitation extremes in term of frequency and magnitude as these region not only showed higher tail risk values in but also showed high changes in tail risk values over time. The findings of these studies match with Yaduvanshi et al. (2021) to some extent, where their study made a similar observation using CMIP5 data and showed that rainfall extremes such as PRCPTOT, R20MM, R95P, R99P, RX1day, and RX5day are projected to rise significantly in tropical, temperate and semi-arid regions of India, under RCP4.5 and RCP8.5 scenario.

A deeper and more comprehensive analysis is necessary to understand the zone-specific predictions of extreme precipitation events, taking into account various factors like air temperature,

precipitable water, moisture flux, and convergence. In terms of the underlying physics guiding the simulations, our reliance is primarily on the methodologies employed by various climate models to replicate these mentioned variables. Nevertheless, it's important to acknowledge that these methodologies may not always be flawless. Therefore, as a standard procedure, we incorporate simulations from multiple models when forming our conclusions.

6.6   Summary and Conclusions

This study evaluates the nature and behavior of risk embedded in the probability distribution of the daily precipitation datasets especially in the tail part. Using the novel concept of the Concentration Profile and Concentration Map, the risk maps are prepared for the observed daily precipitation records and model-simulated daily precipitation records for future time periods over the Indian mainland that provide a comprehensive picture of risk related to extremes over India. This study helps us understand the changes caused by climate change being reflected in the frequency and magnitude of the extremes in the future in a relatively more straightforward manner. Tail risk is evaluated for the past, mainly focusing on the period from 1901 to 2019, which is subdivided into two parts i.e. pre-1970s (1901-1969) and post-1970s (1970-2019) to understand the effect of global climate shift over India. For the future, simulations were obtained for 16 GCMs from CMIP6 for 2020 to 2100, and estimates of tail risk were determined, which were then ensembled to have a multimodel ensemble analysis for different SSP scenarios, namely SSP126, SSP245, SSP370, and SSP585. An incremental change in the tail risk in the future with respect to the historical period indicates both more precipitation days as well as a less uniform distribution of precipitation events with more frequent extremes. Tail risk is our prime concern and accurate assessment based on the data series helps in providing various useful insights. The spatial maps presented in the study illustrate the tail risk on a grid-by-grid, meteorological region-by-region, and Koppen-Geiger Climate zone-by-zone basis, which can be useful for local and regional assessments. The study yields the following explicit conclusions:

(i)   Based on the risk maps for historical periods, a considerable change (increase) in the tail risk is found in the post-1970s with respect to the pre-1970s period. This is true for more than 57% of the grids over India, which can be very well attributed to climate shifts that happened around the mid-to-late 1970s.

(ii)   Similar to the historical period, an increasing trend in tail risk persists in the future time period. All four climate scenarios show an overall increase in the tail risk values (i.e., risk of precipitation extremes), but to different extents relative to the historical time period (i.e., post-1970s). However, a gradual increase in tail risk from SSP126 to SSP245 to SSP70 to SSP585 can be very well observed. Overall, the tail risk for the SSP585 scenario surpasses all other scenarios, indicating its correlation with anthropogenic influences.

(iii)   From the observational analysis of the post-1970 period (with respect to the pre-1970 period), it was noticed that the eastern coast of South India predominantly exhibits an increase in tail risk while the northern and north-western parts of the country showed decreasing tail risk. However, in the future, the western and northern parts of the country show a higher extent of increase in the tail risk than the eastern part of the country

(iv)   The Meteorological subdivision-wise analysis for the historical period revealed that nearly 22 subdivisions out of 34 showed significant changes in the post-1970s period. However, for the future time period, almost all the 34 subdivisions showed significant changes in the tail risk relative to the post-1970s for the climate scenarios of SSP370, and SSP585, respectively.

(v)   The analyses also yielded that subdivisions of Northwestern parts of India (like Gujarat, Vidarbha, Marathwada, West Rajasthan, East Rajasthan, Haryana Chandigarh, and Delhi, Saurashtra, Kutch and Diu, Punjab, Himachal Pradesh), Central part (like West Madhya Pradesh, East Madhya Pradesh, West Uttar Pradesh, Uttaranchal, Chhattisgarh, Orissa) and Southeastern part (like Jharkhand, Coastal Andhra Pradesh, North Interior Karnataka, Rayalseema, Telangana) of India show more increase in tail risks in the future compared to other subdivisions.

(vi)   Climatologically, Tropical (Aw, As), Arid (BWh, BSh), and Temperate (Csa) regions are found to be more susceptible to higher precipitation extremes in terms of frequency and magnitude as these regions not only showed higher tail risk values but also showed high changes in tail risk values over time. This increasing pattern tail risk (i.e., increased frequency and magnitude of extreme precipitation) in these regions will severely impact society as these regions are among the ones that host the major section of the population of the country and are responsible for major agricultural produce.

Assessing and understanding the potential changes in the extreme tail behavior of daily precipitation distributions, encompassing both the intensity and frequency of rainfall extremes, is crucial for informed policy-making at the national or regional level. It is essential to obtain accurate and reliable information on the relative variation in future risks of extreme events. This easily interpretable risk map provided in this study for precipitation extremes considering the geographical variability can be used in risk financing mechanisms for resource allocation by the government. By allocating resources strategically the government agencies associated with the construction or management of water-based infrastructure can minimize the losses and damages associated with climate change. Currently, India has 5334 completed large dams and 411 large dams are under construction (Available on https://cwc.gov.in/sites/default/files/nrld06042019.pdf accessed in July 2023). These dams will experience future changes in climate owing to their long life spans. Hence, the knowledge related to tail risk will enable local decision-makers to effectively address the challenges posed by shifting climate

patterns and develop suitable adaptation strategies in the concerned areas. While our current analysis holds significance, it's important to acknowledge its limitations. Specifically, we have not explored the causes behind the changes in tail risk. Therefore, we recommend future research to investigate regime shifts under different climate scenarios more thoroughly.

SUMMARY AND CONCLUSIONS

This thesis proposes novel and advanced approaches to characterize the tail behavior of the daily precipitation datasets over India by alleviating the limitations associated with the conventional approaches. The following conclusions are drawn from the analyses performed in this thesis.

1. Based on the threshold-based approach introduced by Papalexiou et al. (2013), it is observed that the Lognormal and Pareto type II distributions (both known for their heavy-tailed characteristics) are best suited for fitting the daily precipitation tails over India for both the pre-and post-1970 time periods. This finding underscores the prevalence of heavy-tailed distributions over light-tailed ones, such as the Gamma and Weibull distributions, while modelling precipitation extremes. Furthermore, the research revealed an increase in the tail heaviness of precipitation data from the pre- to post-1970s, indicating the potential influence of global climatic shifts that occurred during the 1970s. Notably, the Gamma distribution consistently underestimates the occurrence and magnitude of extreme events, highlighting its inadequacy in modelling extreme precipitation events within India.

2. The proposed algorithmic approach, based on a novel scalar upper tail indicator known as the 'Obesity Index' (OB), offers advantages in terms of objectivity, robustness, and reproducibility over the traditional threshold-based approaches. The robustness of the proposed approach is established through a simulation study where datasets from four distributions, namely Pareto, Weibull, Lognormal, and Gamma, are generated. Further, the investigations regarding the adequate sample size, as well as the optimum number of random sampling required for the application of the proposed algorithm, are performed. The findings of these investigations suggest that an optimal sample size of 1,000 or greater, along with a minimum of 5,000 random sampling, is sufficient to determine a consistent estimate of OB. Overall, the investigation demonstrates the reliability and effectiveness of the proposed algorithmic approach for upper tail analysis.

3. A comparative analysis is performed between the threshold-based approach (conventional approach) and OB-based approach for effective characterization of the tails of the daily precipitation datasets over India in Chapter 4. All three approaches provided consistent inferences, indicating the presence of heavy-tailed behavior in approximately 78.21% of grid locations across India. At a few grids over India, the OB-based approach is noticed to yield lighter tails, whereas the threshold-based methods tend to suggest exponential tails. It is very important to acknowledge that while these approaches generally align in identifying heavy-tailed behavior, they may diverge in their assessments due to the

inherent differences in their methodologies. Overall, the analysis highlights the suitability of heavy-tailed distributions for modeling daily precipitation extremes over India. The OB-based approach stands out for its ability to provide a quantitative measure for assessing the tail behavior in daily precipitation data without the need for assuming specific threshold values to categorize tails. This feature enhances the precision and versatility of tail analysis in hydroclimatology and presents it as a good alternative diagnostic tool for assessing tail behavior.

4. Characterization of the tail behavior in the projections of daily precipitation datasets for future time periods reveals a notable increase in the frequency of occurrence of extreme precipitation events, particularly under the RCP8.5 scenario compared to the RCP4.5 scenario for simulated datasets from the Coordinated Regional Climate Downscaling Experiment (CORDEX). The notable rise in the frequency of extreme precipitation events under the RCP8.5 scenario is primarily attributed to the anthropogenic forcings, driven by higher greenhouse gas emissions and greater radiative forcings, resulting in rising temperatures. These findings highlight the broader impacts of climate change on precipitation extremes over India. Additionally, the analysis of OB in relation to the topography of India unveiled a non-linear decrease in OB with elevations up to 2500 meters, followed by an increase beyond this elevation. Importantly, this elevation-related variation in OB is found to be spatially correlated with the Köppen-Geiger climate classification system of India. This correlation suggests that the climate type and elevation have a discernible influence on the tail heaviness of precipitation datasets, further emphasizing the complex interplay of factors shaping the extreme precipitation patterns over India.

5. Failing to select the appropriate class/family of distributions that can provide the best fit to the upper tail of the dataset before selecting a model can lead to either underestimation or overestimation of the quantiles. In an effort to address this issue, a comprehensive Decision support system (DSS) is proposed in Chapter 5 of this thesis, which can characterize the tails of the probability distributions into various families. The conventional DSSs have limitations in (i) efficient segregation of Lognormal distribution from the regularly varying and subexponential distribution families and (ii) identifying the distributions from hyper-exponential class. Advanced graphical tools such as Concentration Profile, concentration adjusted expected shortfall, discriminant moment ratio plot, maximum to sum plot, and Zenga plot can effectively identify different distributions simultaneously if used in an appropriate order. This formed the basis for proposing the comprehensive DSS in this thesis. The proposed comprehensive DSS is found to be more capable and robust than the conventional DSS in respect of its ability to categorize the tails of distributions into Classes B\A (Pareto type), C\B (regularly varying), D\C (subexponential), E (Exponential type), hyper-exponential type (outside class E). The robustness of the proposed comprehensive DSS over the conventional DSSs is established through a simulation study by generating datasets having various sample sizes from various representative probability distributions belonging to various classes and reassigning them successfully to their respective classes. The percentage of success in the

characterization of distributions in various classes is found to be appreciable at smaller sample sizes (<1000) and quite high for sample sizes (>1000). This indicates the validity of the proposed approach for the characterization of the tails of the precipitation datasets.

6. The application of the proposed DSS for characterizing the tails of the daily precipitation over India revealed a successful categorization into 4 classes, namely class C\B, D\C E, and limiting case LN. Results indicate that out of 4801 grids, 63 (1.31%), 3671 (76.46%), 106 (2.21%), and 961 (20.01%) grids belong to classes C\B, D\C, E, and LN distribution, respectively. The findings from the study revealed that around 98% of the daily precipitation records are found to exhibit subexponential tails. Overall, it is concluded that the probability distribution tails for daily precipitation are heavy-tailed, and the spatial information on these classes of distributions over India could be used as a priori information for regional studies to identify suitable probability distribution.

7. The development of a framework for assessing the embedded risk associated with the tails of the daily precipitation is performed based on the concept of the Concentration Profile (CP) and Concentration Map (CM). The graphical framework of CM presented in Chapter 7 of this thesis is intended to provide useful insights into the tail risk of the precipitation datasets and to compare the distribution of datasets having different scales and magnitudes as the proposed framework is an easy and scale-free approach which is the major advantage associated with CM.

8. The application of the proposed framework revealed a considerable increase at more than 57% of the grids over India in the tail risk in the post-1970s relative to the pre-1970s period. Results are very well attributed to climate shifts that happened around the mid-to-late 1970s. Similar to the historical period, the increasing trend in tail risk persists in the future. All four climate scenarios show an overall increase in the tail risk values (i.e., risk of precipitation extremes), but to different extents with respect to the historical (Relative changes%) and spatial variation. However, a gradual increase in tail risk from SSP126 to SSP245 to SSP70 to SSP585 scenarios for simulations from 16 General Circulation Models (GCMs) participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6) can very well be observed. Overall, SSP585 surpasses all other scenarios in terms of tail risk, indicating that such an increase in future precipitation extremes is especially due to anthropogenic influences.

9. Climatologically, Tropical (Aw, As), Arid (BWh, BSh), and Temperate (Csa) regions are found to be more susceptible to higher precipitation extremes in terms of frequency and magnitude as these regions not only showed higher tail risk values but also showed high changes in tail risk values over time. This increasing pattern tail risk (i.e., increased frequency and magnitude of extreme precipitation) in these regions will severely impact society as these regions are among the ones that host the major section of the population of the country and are responsible for major agricultural produce.

In summary, the approaches proposed in this thesis alleviate most of the limitations associated with the conventional approaches to characterize daily precipitation tails. From a practical standpoint, these approaches are straightforward, user-friendly, and have the potential for wide application across diverse hydroclimatological regions of the globe.

7.1   Scope for Future Work

The present research is devoted to understanding the frequency of occurrence of daily precipitation extremes through the assessment of tail behavior. This thesis describes novel and comprehensive ways to assess tail behavior and estimate the embedded risks associated with it. However, there is still considerable scope for extending the research work presented in this thesis. Future work should explore the effectiveness of the proposed approaches on various other hydroclimatic regions in various parts of the world to strengthen the conclusions drawn based on the present study. Future scope exists for:

1. Investigating the causes of heavy tails in precipitation extremes within various parts of India through detection and attribution. Understanding the underlying mechanisms behind heavy tails in different regions is crucial in the current scenario of changing climate.

2. Exploring the relationship between different physical processes and corresponding tail behavior

3. Exploring the potential of other upper tail indicators in the characterization of tail behavior, especially when the data lengths are limited

4. Developing an automated algorithmic mechanism to discriminate between distributions belonging to the Classes C\B and D\C by devising efficient procedures for examining the linearity of log-log plots based on estimated correlation coefficient values, similar to the Mean Excess Function (MEF) proposed by Nerantzaki and Papalexiou (2019). This can automate the analysis and alleviate the limitation of the present DSS by providing an efficient algorithmic method for model selection and decision-making for practicing engineers.

5. Developing frameworks utilizing different threshold-based, graphical, and scalar approaches that can incorporate the non-stationarity associated with hydroclimatic datasets, while assessing the tail behavior.

6. Exploring the variations in tail behavior of daily precipitation over India by using different global precipitation products as well as by considering various climate classifications apart from the Köppen Geiger classification used in this thesis. This investigation can help in addressing the ambiguity associated with different datasets. Additionally, the remarkably extensive observational dataset, characterized by a substantial number of stations, extended time series, and numerous explanatory

variables, provides a unique opportunity to gain deeper insights into the upper tail behavior of precipitation over India—an endeavor rarely feasible with conventional observational datasets.

7. Understanding the propagation of risks from extreme precipitations to subsequent floods within the Indian watersheds with the help of the tail characterization frameworks proposed in this thesis.

8. Application of the risk assessment tool in scenarios such as PMP (Probable Maximum Precipitation) estimation, flood assessment of dams, or other related studies

9. Developing a framework for integrating the tail risks of daily precipitation in hazard evaluations by considering factors such as exposure to population, human establishments, etc.

REFERENCES

- Aadhar, S., & Mishra, V. (2020). On the projected decline in droughts over South Asia in CMIP6 multimodel ensemble. *Journal of Geophysical Research: Atmospheres*, 125(20), e2020JD033587.

- Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of banking & finance*, *26*(7), 1487-1503.

- Adams, B. J., Fraser, H. G., Howard, C. D., & Sami Hanafy, M. (1986). Meteorological data analysis for drainage system design. *Journal of environmental Engineering*, *112*(5), 827-848.

- AghaKouchak, A., & Nasrollahi, N. (2010). Semi-parametric and parametric inference of extreme value models for rainfall data. *Water resources management*, *24*, 1229-1249.

- AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O. & Sadegh, M. (2020). Climate extremes and compound hazards in a warming world. *Annual Review of Earth and Planetary Sciences*, *48*, 519-548.

- Ajayamohan, R. S., & Rao, S. A. (2008). Indian Ocean dipole modulates the number of extreme rainfall events over India in a warming environment. *Journal of the Meteorological Society of Japan. Ser. II*, *86*(1), 245-252. doi:10.2151/jmsj.86.245.

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705.

- Aksoy, H. (2000). Use of gamma distribution in hydrological analysis. *Turkish Journal of Engineering and Environmental Sciences*, *24*(6), 419-428.

- Al-Zahrani, M., & Husain, T. (1998). An algorithm for designing a precipitation network in the south-western region of Saudi Arabia. *Journal of Hydrology*, *205*(3-4), 205-216.

- Ali, H., & Mishra, V. (2018). Increase in subdaily precipitation extremes in India under 1.5 and 2.0 C warming worlds. *Geophysical Research Letters*, *45*(14), 6972-6982.

- Anandhi, A., Frei, A., Pierson, D.C., Schneiderman, E.M., Zion, M.S., Lounsbury, D., & Matonse, A.D. (2011) Examination of change factor methodologies for climate change impact assessment. *Water Resources Research*, 47, W03501. https://doi.org/10.1029/2010WR009104.

- Arguez, A., & Vose, R. S. (2011). The definition of the standard WMO climate normal: The key to deriving alternative climate normals. *Bulletin of the American Meteorological Society*, *92*(6), 699-704. doi:10.1175/2010BAMS2955.1.

- Arnold, B. C. (2012). On the Amato inequality index. *Statistics & Probability Letters*, *82*(8), 1504-1506.

- Babu, V. B., & Hooda, B. K. (2018). Fuzzy majority approach for modeling spatial and temporal distributions of daily rainfall in western zone of Haryana. *Int. J. Agricult. Stat. Sci*, *14*(1), 57-67.

- Balkema, A. A., & De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, *2*(5), 792-804.

- Baines, P. G., & Folland, C. K. (2007). Evidence for a rapid global climate shift across the late 1960s. *Journal of Climate*, 20(12), 2721-2744.

- Barucci, E., & Fontana, C. (2003). *Financial markets theory*. London: Springer-Verlag.

- Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., & Wood, E. F., (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modelling. *Hydrology and Earth System Sciences*, 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017.

- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data*, *5*(1), 1-12.

- Bernardara, P., Schertzer, D., Sauquet, E., Tchiguirinskaia, I., & Lang, M. (2008). The flood probability distribution tail: how heavy is it?. *Stochastic Environmental Research and Risk Assessment*, *22*, 107-122.

- Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., De Waal, D., & Ferro, C., (2004). *Statistics of extremes: Theory and applications, Statistics of Extremes: Theory and Applications*. Wiley. https://doi.org/10.1002/0470012382.

- Beguería, S. (2005). Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology*, *303*(1-4), 215-230.

- Ben-Zvi, A. (2009). Rainfall intensity–duration–frequency relationships derived from large partial duration series. *Journal of Hydrology*, *367*(1-2), 104-114. doi:10.1016/j.jhydrol.2009.01.007.

- Beskow, S., Caldeira, T. L., de Mello, C. R., Faria, L. C., & Guedes, H. A. S. (2015). Multiparameter probability distributions for heavy rainfall modeling in extreme southern Brazil. *Journal of Hydrology: Regional Studies*, Volume 4, Part B, 123-133, ISSN 2214-5818. https://doi.org/10.1016/j.ejrh.2015.06.007.

- Bhavana, C., Munirajappa, R., Surendra, H., & Rathod, S. (2012). Modeling of daily rainfall using gamma probability distribution. *Environment and Ecology*, *30*(3B), 884-887.

- Bhattacharya, B., Mohanty, S., & Singh, C. (2022). Assessment of the potential of CMIP6 models in simulating the sea surface temperature variability over the tropical Indian Ocean. *Theoretical and Applied Climatology*, 148(1-2), 585-602.

- Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1987). *Regular variation* (No. 27). Cambridge university press..

- Biondi, F., Gershunov, A., & Cayan, D. R. (2001). North Pacific decadal climate variability since 1661. *Journal of climate*, *14*(1), 5-10. https:// doi.org/10.1175/1520-0442(2001)014,0005:NPDCVS.2.0.CO

- Bisht, D. S., Chatterjee, C., Raghuwanshi, N. S., & Sridhar, V. (2018a). Spatio-temporal trends of rainfall across Indian river basins. *Theoretical and applied climatology*, *132*, 419-436.

- Bisht, D. S., Chatterjee, C., Raghuwanshi, N. S., & Sridhar, V. (2018b). An analysis of precipitation climatology over Indian urban agglomeration. *Theoretical and Applied Climatology*, *133*, 421-436. https://doi.org/10.1007/s00704-017-2200-z.

- Brazauskas, V., & Kleefeld, A. (2016). Modeling severity and measuring tail risk of Norwegian fire claims. *North American Actuarial Journal*, *20*(1), 1-16.

- Bryson, M. (1974). Heavy-Tailed Distributions: Properties and Tests. *Technometrics*, *16(1)*, 61-68. Doi:10.2307/1267493

- Caires, S. (2009). A comparative simulation study of the annual maxima and the peaks-over-threshold methods. *Deltares report 1200264-002 for Rijkswaterstaat, Waterdienst*.

- Cavanaugh, N. R., Gershunov, A., Panorska, A. K., & Kozubowski, T. J. (2015). The probability distribution of intense daily precipitation. *Geophysical Research Letters*, *42*(5), 1560-1567. doi:10.1002/2015GL063238.

- Caeiro, F., Gomes, M.I., Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat – Statistical Journal*, *3*, 113-136.

- Changyong, F. E. N. G., Hongyue, W. A. N. G., Naiji, L. U., Tian, C. H. E. N., Hua, H. E., & Ying, L. U. (2014). Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry*, *26*(2), 105.

- Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, *63*(250), 318-351.

- Chowdary, J. S., Gnanaseelan, C., Vaid, B. H., & Salvekar, P. S. (2006). Changing trends in the tropical Indian Ocean SST during La Nina years. *Geophysical research letters*, *33*(18).https://doi.org/10.1029/2006GL026707.

- Chen, J., & Brissette, F. P. (2014). Stochastic generation of daily precipitation amounts: review and evaluation of different models. *Climate Research*, *59*(3), 189-206. Doi:10.3354/cr01214

- Chavan, S. R., & Srinivas, V. V. (2021). Evaluation of three approaches to probable maximum precipitation estimation: a study on two Indian river basins. *Theoretical and Applied Climatology*, *144*, 731-749.

- Choudhary, A., & Dimri, A. P. (2019). On bias correction of summer monsoon precipitation over India from CORDEX-SA simulations. *International Journal of Climatology*, *39*(3), 1388-1403.

- Choudhury, B. A., Rajesh, P. V., Zahan, Y., & Goswami, B. N. (2021). Evolution of the Indian summer monsoon rainfall simulations from CMIP3 to CMIP6 models. *Climate Dynamics*, 1-26.

- Chen, J., Brissette, F.P., Chaumont, D., & Braun, M. (2013). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research*, *49*, 4187–4205. Doi:10.1002/wrcr.20331.

- Chen, S., Yu, B., & Chen, W. (2015). An interdecadal change in the influence of the spring Arctic Oscillation on the subsequent ENSO around the early 1970s. *Climate Dynamics*, *44*, 1109-1126. https://doi.org/10.1007/s00382-014-2152-2.

- Chistyakov, V.P. (1964) A theorem on sums of independent positive random variables and its applications to branching process. *Theory of Probability and its Application*, *9*, 640-648.

- Chow, V. T. (1964). Handbook of applied hydrology: a compendium of water-resources technology. In *Handbook of applied hydrology: a compendium of water-resources technology* (pp. 1525-1525).

- Chen, K., & Cheng, T. (2022). Measuring tail risks. *The Journal of Finance and Data Science*, *8*, 296-308.

- Chavan, S.R. & Srinivas, V.V. (2015). Effect of DEM Source on Equivalent Horton Strahler Ratio based GIUH for Catchments in Two Indian River Basins. *Journal of Hydrology*, Elsevier, Netherlands, 528, Issues 1-4, pp.463-489. Doi: 10.1016/j.jhydrol.2015.06.049.

- Choudhary, A., Dimri, A. P., & Maharana, P. (2017). Assessment of CORDEX-SA experiments in representing precipitation climatology of summer monsoon over India. *Theoretical and Applied Climatology*, *134*, 283-307.

- Christensen, J. H., Boberg, F., Christensen, O. B., & Lucas-Picher, P. (2008). On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical research letters*, *35*(20). Doi:10.1029/2008GL035694.

- Chowdary, J. S., Gnanaseelan, C., Vaid, B. H., & Salvekar, P. S. (2006). Changing trends in the tropical Indian Ocean SST during La Nina years. *Geophysical research letters*, *33*(18).

- Cirillo, P. (2013). Are your data really Pareto distributed?. *Physica A: Statistical Mechanics and its Applications*, *392*(23), 5947-5962.

- Cirillo, P., & Taleb, N. N. (2016a). On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and its Applications*, *452*, 29-45.

- Cirillo, P., & Taleb, N. N. (2016b). Expected shortfall estimation for apparently infinite-mean models of operational risk. *Quantitative Finance*, *16*(10), 1485-1494.

- Cirillo, P., & Taleb, N. N. (2020). Tail risk of contagious diseases. *Nature Physics, 16*(6):606-613.

- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661-703.

- Cooke, R. M., & Nieboer, D. (2011). Heavy-tailed distributions: Data, diagnostics, and new developments. *Resources for the Future Discussion Paper*, (11-19), Washington. https://doi.org/10.2139/ssrn.1811043.

- Cooke, R. M., Nieboer, D., & Misiewicz, J. (2014). *Fat-Tailed Distributions: Data, Diagnostics and Dependence, Volume 1* (Vol. 1). John Wiley & Sons, Inc., Hoboken, NJ.

- Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208, p. 208). London: Springer.

- Coles, S., Pericchi, L. R., & Sisson, S. (2003). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, *273*(1-4), 35-50.

- Cooke, R. M., Nieboer, D., & Misiewicz, J. (2014). *Fat-Tailed Distributions: Data, Diagnostics and Dependence, Volume 1* (Vol. 1). John Wiley & Sons.

- Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *Weather, Climate, and Society*, *2*(1), 60-68.

- Cobb, C. W., & Douglas, P. H. (1928). A theory of production.

- Curriero, F. C., Patz, J. A., Rose, J. B., & Lele, S. (2001). The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *American journal of public health*, *91*(8), 1194-1199.

- Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of hydrology*, *18*(3-4), 257-271.

- Dai, T., Dong, W., Guo, Y., Hong, T., Ji, D., Yang, S., ... & Zhu, X. (2018). Understanding the abrupt climate change in the mid-1970s from a phase-space transform perspective. *Journal of Applied Meteorology and Climatology*, *57*(11), 2551-2560. https://doi.org/10.1175/JAMC-D-17-0345.1.

- Davies, A. S., Hernebring, C., Svensson, G., & Gustafsson, L. G. (2008). The impacts of climate change and urbanisation on drainage in Helsingborg, Sweden: Suburban stormwater. *Journal of Hydrology*, *350* (1-2), 114-125. doi:10.1016/j.jhydrol.2007.11.006.

- Dash, S., & Maity, R. (2019). Temporal evolution of precipitation-based climate change indices across India: contrast between pre-and post-1975 features. *Theoretical and Applied Climatology*, *138*(3-4), 1667-1678. https://doi.org/10.1007/s00704-019-02923-8.

- Dash, S., Kulkarni, M.A., Mohanty, U., & Prasad, K. (2009). Changes in the characteristics of rain events in India. *Journal of Geophysical Research: Atmosphere,* 114, D10.

- Das, B., & Ghosh, S. (2016). Detecting tail behavior: mean excess plots with confidence bounds. *Extremes*, *19*, 325–349. https://doi.org/10.1007/s10687-015-0238-9

- Das, S., Sarkar, S., & Kanungo, D. P. (2022). Rainfall-induced landslide (RFIL) disaster in Dima Hasao, Assam, Northeast India. https://doi.org/10.1007/s10346-022-01962-z.

- Dash, S. K., Kulkarni, M. A., Mohanty, U. C., & Prasad, K. (2009). Changes in the characteristics of rain events in India. *Journal of Geophysical Research: Atmospheres*, *114*(D10).

- Daníelsson, J., Haan, de L., Peng, L., & Vries, de C.G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, *76*, 226–248.

- Danielsson, J., Jorgensen, B. N., Sarma, M., & Vries, de, C. G. (2006). Comparing downside risk measures for heavy tailed distributions. *Economics letters*, *92*(2), 202-208.

- Daníelsson, J., Ergun, L.M., Haan, de L., & Vries, de C.G. (2019). *Tail Index Estimation: Quantile Driven Threshold Selection.* Staff Working Papers 19-46, Bank of Canada.

- Dave, H., & James, M. E. (2017). Characteristics of intense rainfall over Gujarat State (India) based on percentile criteria, *Hydrological Sciences Journal*, *62*(12), 2035-2048. DOI: 10.1080/02626667.2017.1357818.

- De Michele, C., & Avanzi, F. (2018). Superstatistical distribution of daily precipitation extremes: A worldwide assessment. *Scientific reports*, *8*(1), 14204.

- De Michele, C. (2019). Advances in deriving the exact distribution of maximum annual daily precipitation. *Water*, *11*(11), 2322.

- De Sousa, B., & Michailidis, G. (2004). A diagnostic plot for estimating the tail index of a distribution. *Journal of Computational and Graphical Statistics*, *13*(4), 974-995.

- DeNicola, E., Aburizaiza, O. S., Siddique, A., Khwaja, H., & Carpenter, D. O. (2015). Climate change and water scarcity: The case of Saudi Arabia. *Annals of global health*, *81*(3), 342-353.

- Deepthi, B., & Sivakumar, B. (2022). General circulation models for rainfall simulations: Performance assessment using complex networks. *Atmospheric Research*, *278*, 106333.

- Dikshit, A., Sarkar, R., Pradhan, B., Segoni, S., & Alamri, A. M. (2020). Rainfall induced landslide studies in Indian Himalayan region: a critical review. *Applied Sciences*, *10*(7), 2466.

- Diakakis, M. (2020). Types of behavior of flood victims around floodwaters. Correlation with situational and demographic factors. *Sustainability*, *12*(11), 4409.

- Dutta, R., & Maity, R. (2022). Value addition in coupled model intercomparison project phase 6 over phase 5: global perspectives of precipitation, temperature and soil moisture fields. *Acta Geophysica*, *70*(3), 1401-1415.

- Edwards, W. (1992). *Utility Theories: Measurements and Applications*. Springer

- Efron, B. (1979). Bootstrap methods: Another look at jackknife. *Annals of Statistics*, *7*, 1-26.

- Ehsanzadeh, E., El Adlouni, S., & Bobée, B. (2010). Frequency analysis incorporating a decision support system for hydroclimatic variables. *Journal of Hydrologic Engineering*, *15*(11), 869-881.

- El Adlouni, S., Bobée, B., & Ouarda, T. B. (2008). On the tails of extreme event distributions in hydrology. *Journal of hydrology*, *355*(1-4), 16-33. doi:10.1016/J.JHYDROL.2008.02.011

- Eliazar, I. I., & Sokolov, I. M. (2010). Gini characterization of extreme-value statistics. Physica A: Statistical Mechanics and its Applications, 389(21), 4462-4472.

- Eliazar, I. (2018). A tour of inequality. *Annals of Physics*, *389*, 306-332.

- Embrechts, P. & Goldie, C. (1980). On Closure and factorization properties of subexponential and related distributions. *Journal of Australian Mathematical Society. Series A. Pure Mathematics and Statistics,* 29(2), 243-256.doi:10.1017/S1446788700021224

- Embrechts, P., & Goldie, C. M. (1982). On convolution tails. *Stochastic Processes and their Applications*, *13*(3), 263-278.

- Embrechts, P., Klüppelberg, C., Thomas, M. (1997) Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling extremal events: for insurance and finance* (Vol. 33). Springer Science & Business Media.

- Embrechts, P. (2003). Extremes in economics and the economics of extremes. *Extreme values in finance, telecommunications, and the environment*, 169-83. Chapman and Hall/CRC.

- Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling extremal events: for insurance and finance* (Vol. 33). Springer Science & Business Media.

- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937-1958.

- Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y.

- Farooq, M., Shafique, M., & Khattak, M. S. (2018). Flood frequency analysis of river swat using Log Pearson type 3, Generalized Extreme Value, Normal, and Gumbel Max distribution methods. *Arabian Journal of Geosciences*, *11*, 1-10.

- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. Vol. 2, J. Willey and Sons, New York, 766.

- Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society* (Vol. 24, No. 2, pp. 180-190). Cambridge University Press. https://doi.org/10.1017/S0305004100015681.

- Foss, S., Korshunov, D., & Zachary, S. (2013). *An introduction to heavy-tailed and subexponential distributions* (Vol. 6, pp. 0090-6778). New York: Springer. doi: 10.1007/978-1-4419-9473-8.

- Fontanari, A., Cirillo, P., & Oosterlee, C. W. (2018a). From concentration profiles to concentration maps. New tools for the study of loss distributions. *Insurance: Mathematics and Economics*, *78*, 13-29.

- Fontanari, A., Taleb, N. N., & Cirillo, P. (2018b). Gini estimation under infinite variance. *Physica A: Statistical Mechanics and its Applications*, *502*, 256-269.

- Fontanari, A. (2019). *Lorenz-based quantitative risk management.* Doctoral dissertation, Delft University of Technology. https://doi.org/10.4233/uuid:0c5b50a5-4514-431d-a31a-b1f4ae2c0713.

- Furman, E., Wang, R., & Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking & Finance*, *83*, 70-84.

- Geiger, R. (1954). Landolt-Börnstein–Zahlenwerte und FunktionenausPhysik, Chemie, Astronomie, Geophysik und Technik, alteSerie Vol. 3. *Ch. Klassifikation der Klimatenach W. Köppen.–Springer, Berlin*, 603-607.

- Ghosh, S. and Resnick, S. (2010) A discussion on mean excess plots. *Stochastic Processes and their Applications*, Volume 120, Issue 8, pages 1492-1517, ISSN 0304-4149. https://doi.org/10.1016/j.spa.2010.04.002.

- Ghosh, S., Vittal, H., Sharma, T., Karmakar, S., Kasiviswanathan, K. S., Dhanesh, Y., ... & Gunthe, S. S. (2016). Indian summer monsoon rainfall: implications of contrasting trends in the spatial variability of means and extremes. *PloS one*, *11*(7), e0158670. doi:10.1371/journal.pone.0158670.

- Ghosh, S., Das, D., Kao, S. C., & Ganguly, A. R. (2012). Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. *Nature Climate Change*, *2*(2), 86-91. doi:10.1038/nclimate1327.

- Ghosh, S. & Mujumdar, P.P. (2009). Climate change impact assessment: Uncertainty modeling with imprecise probability. *Journal of Geophysical Research*, 114. 10.1029/2008JD011648.

- Gini, C. (1912). Variabilità E Mutabilità. Reprinted in: Variabilità e Mutabilità, E Pizetti and T Salvemini, Memorie di MetodologicaStatistica, LibreriaEredi Virgilio Veschi, Rome.

- Giorgi, F., Im, E. S., Coppola, E., Diffenbaugh, N. S., Gao, X. J., Mariotti, L., & Shi, Y. (2011). Higher hydroclimatic intensity with global warming. *Journal of Climate*, *24*(20), 5309-5324. doi:10.1175/4092011JCLI3979.1.

- Giorgi, F., Coppola, E., & Raffaele, F. (2014). A consistent picture of the hydroclimatic response to global warming from multiple indices: Models and observations. *Journal of Geophysical Research: Atmospheres*, *119*(20), 11-695. https://doi.org/10.1002/2014JD022238

- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, 423-453. https://doi.org/10.2307/1968974.

- Goldie, C. M., & Klüppelberg, C. (1998). Subexponential distributions. *A practical guide to heavy tails: statistical techniques and applications*, 435-459. Cambridge, MA: Birkhauser Boston Inc.

- Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning (p. 126). New York: Addison-Wesley. Reading, Mass.

- Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., & Xavier, P. K. (2006). Increasing trend of extreme rain events over India in a warming environment. *Science*, *314*(5804), 1442-1445.

- Groisman, P. Y., Karl, T. R., Easterling, D. R., Knight, R. W., Jamason, P. F., Hennessy, K. J., ... & Zhai, P. M. (1999). Changes in the probability of heavy precipitation: important indicators of climatic change. *Weather and climate extremes: changes, variations and a perspective from the insurance industry*, 243-283. Springer, Dordrecht.

- Graham, N. E. (1994). Decadal-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: Observations and model results. *Climate Dynamics*, *10*, 135-162.

- Greselin, F., Pellegrino, S., & Vernizzi, A. (2017). *Lorenz versus Zenga Inequality Curves: a New Approach to Measuring Tax Redistribution and Progressivity* (No. 046).

- Guhathakurta, P., & Rajeevan, M. (2008). Trends in the rainfall pattern over India. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *28*(11), 1453-1469.

- Guhathakurta, P., Sreejith, O. P., & Menon, P. A. (2011). Impact of climate change on extreme rainfall events and flood risk in India. *Journal of earth system science*, *120*, 359-373.

- Gu, X., Zhang, Q., Singh, V. P., Liu, L., & Shi, P. (2017). Spatiotemporal patterns of annual and seasonal precipitation extreme distributions across China and potential impact of tropical cyclones. *International Journal of Climatology*, *37*(10), 3949-3962.

- Gusain, A., Ghosh, S., & Karmakar, S. (2020). Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmospheric Research*, *232*, 104680.

- Gunwani, P., & Mohan, M. (2017). Sensitivity of WRF model estimates to various PBL parameterizations in different climatic zones over India. *Atmospheric research*, *194*, 43-65.

- Gupta, S.K. (2011). *Modern Hydrology and Sustainable Water Development*. John Wiley & Sons. https://doi.org/10.1002/9781444323962.

- Gupta, V., Singh, V., & Jain, M. K. (2020). Assessment of precipitation extremes in India during the 21st century under SSP1-1.9 mitigation scenarios of CMIP6 GCMs. *Journal of Hydrology*, 590, 125422.

- Guhathakurta, P. & Rajeevan, M. (2008). Trends in the rainfall pattern over India. *International Journal of Climatology*, 28. 1453 – 1469. 10.1002/joc.1640.

- Gu, X., Zhang, Q., Singh, V. P., Liu, L., & Shi, P. (2017). Spatiotemporal patterns of annual and seasonal precipitation extreme distributions across China and potential impact of tropical cyclones. *International Journal of Climatology*, *37*(10), 3949-3962.

- Gupta, N., & Chavan, S. R. (2021). Assessment of temporal change in the tails of probability distribution of daily precipitation over India due to climatic shift in the 1970s. *Journal of Water and Climate Change*, *12*(6), 2753-2773.

- Gupta, N., & Chavan, S. R. (2022). Characterizing the tail behaviour of daily precipitation probability distributions over India using the obesity index. *International Journal of Climatology*, *42*(4), 2543-2565.

- Gupta, N., & Chavan, S. R. (2023a). Investigating the tail behaviour and associated risk with daily discharges in South Indian Rivers. *Stochastic Environmental Research and Risk Assessment*, 1-17.

- Gupta, N. & Chavan, S. R. (2023b). A comprehensive decision support system for characterization of tails of probability distributions of daily precipitation. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2023.130282.

- Gusain, A., Vittal, H., Kulkarni, S., Ghosh, S., & Karmakar, S. (2019). Role of vertical velocity in improving finer scale statistical downscaling for projection of extreme precipitation. *Theoretical and Applied Climatology*, *137*, 791-804. https://doi.org/10.1007/s00704-018-2615-1.

- Haddad, K., & Rahman, A. (2011). Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia. *Stochastic Environmental Research and Risk Assessment*, *25*, 415-428. https://doi.org/10.1007/s00477-010-0412-1.

- Hassani, A. (2012). Applications of Cobb-Douglas Production Function in Construction Time-Cost Analysis.

- Heo, J. H., Salas, J. D., & Boes, D. C. (2001). Regional flood frequency analysis based on a Weibull model: Part 2. Simulations and applications. *Journal of hydrology*, *242*(3-4), 171-182.

- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3, 1163-1174.

- Hill, S. A. (2019). A measure for characterizing heavy-tailed networks. *arXiv preprint arXiv:1907.04808*.

- Hobbi, S. (2021). *Global characteristics of extreme precipitation and variation of climate types from Köppen-Geiger classification using different datasets* (Doctoral dissertation, University of Saskatchewan).

- Hobbi, S., Papalexiou, S. M., Rajulapati, C. R., Nerantzaki, S. D., Markonis, Y., Tang, G., & Clark, M. P. (2022). Detailed investigation of discrepancies in Köppen-Geiger climate classification using seven global gridded products. *Journal of Hydrology*, *612*, 128121.

- Hosking, J. R. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *52*(1), 105-124.

- Hsu, P. C. (2016). Global monsoon in a changing climate. The Monsoons and Climate Change: Observations and Modeling, LM Véspoli de Carvalho and C. Jones, Eds., Springer, 7–24.

- Huang, J., Yu, H., Dai, A., Wei, Y., & Kang, L. (2017a). Drylands face potential threat under 2 C global warming target. *Nature Climate Change*, *7*(6), 417-422, https://doi.org/10.1038/nclimate3275.

- Huang, J., Li, Y., Fu, C., Chen, F., Fu, Q., Dai, A., ... & Wang, G. (2017b). Dryland climate change: Recent progress and challenges. *Reviews of Geophysics*, *55*(3), 719-778.

- Huang, J., Mondal, S. K., Zhai, J., Fischer, T., Wang, Y., Su, B., ... & Jiang, T. (2022). Intensity-area-duration-based drought analysis under 1.5 C–4.0 C warming using CMIP6 over a climate hotspot in South Asia. *Journal of Cleaner Production*, *345*, 131106.

- Hussain, T., Bakouch, H. S., & Chesneau, C. (2019). A new probability model with application to heavy-tailed hydrological data. *Environmental and Ecological Statistics*, *26*, 127-151.

- Hull, J. (2012). *Risk management and financial institutions,+ Web Site* (Vol. 733). John Wiley & Sons.

- IPCC 2013 Climate Change (2013). The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, et al., editors Cambridge, United Kingdom and New York, NY, USA. Cambridge University Press.

- Jaiswal, R., Mall, R. K., Singh, N., Lakshmi Kumar, T. V., & Niyogi, D. (2022). Evaluation of bias correction methods for regional climate models: Downscaled rainfall analysis over diverse agroclimatic zones of India. *Earth and Space Science*, 9(2), e2021EA001981.

- Jawitz, J. W., & Mitchell, J. (2011). Temporal inequality in catchment discharge and solute export. *Water Resources Research*, *47*(10), 1–16. doi:10.1029/2010WR010197.

- Jacob, D. A. N. I. E. L. A., & Hagemann, S. T. E. F. A. N. (2007). Intensification of the hydrological cycle: an important signal of climate change. In *Global change: Enough water for all?* (pp. 170-173). Wissenschaftliche Auswertungen.

- Jacques-Coper, M., & Garreaud, R. D. (2015). Characterization of the 1970s climate shift in South America. *International Journal of Climatology*, *35*(8), 2164-2179. https://doi.org/10.1002/joc.4120.

- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158-171. https://doi.org/10.1002/qj.49708134804.

- Johnson, N. I., & Kotz, S., (1970). *Continuous Univariate Distributions*. 1–2. New York: Wiley.

- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions, volume 2* (Vol. 289). John wiley & sons.

- Jorion, P. (2007). *Value at risk: the new benchmark for managing financial risk*. The McGraw-Hill Companies, Inc.

- Jones, B. L., & Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, *7*(4), 44-54.

- Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, *25*(8-12), 1287-1304. doi:10.1016/S0309-1708(02)00056-8.

- Katz, R. W. (2010). Statistics of extremes in climate change. *Climatic change*, *100*(1), 71-76.

- Kamruzzaman, M., Shahid, S., Islam, A. T., Hwang, S., Cho, J., Zaman, M. A. U., ...& Hossain, M. B. (2021). Comparison of CMIP6 and CMIP5 model performance in simulating historical precipitation and temperature in Bangladesh: a preliminary study. *Theoretical and Applied Climatology*, 145, 1385-1406.

- Kiran, K. G., & Srinivas, V. V. (2021). A Mahalanobis distance-based automatic threshold selection method for peaks over threshold model. *Water Resources Research*, *57*(1), e2020WR027534.

- Klüppelberg, C. (1988). Subexponential distributions and integrated tails. *Journal of Applied Probability*, *25*(1), 132-141. https://doi.org/10.2307/3214240.

- Kotz, S., & Nadarajah, S. (2000). *Extreme value distributions: theory and applications*. world scientific. Imperial College Press, London. https://doi.org/10.1142/p191.

- Koutsoyiannis, D. (2004a). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. *Hydrological sciences journal*, *49*(4).

- Koutsoyiannis, D. (2004b). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Hydrological Sciences Journal*, *49*(4).

- Koutsoyiannis, D. (2008). Probability and statistics for geophysical processes. *Athens: National Technical University of Athens.* https://doi.org/10.13140/RG.2.1.2300.1849/1.

- Kozubowski, T. J., Panorska, A. K., Qeadan, F., Gershunov, A., & Rominger, D. (2008). Testing exponentiality versus Pareto distribution via likelihood ratio. *Communications in Statistics-Simulation and Computation*, *38*(1), 118-139. doi:10.1080/03610910802439121.

- Konapala, G., Mishra, A., & Leung, L. R. (2017). Changes in temporal variability of precipitation over land due to anthropogenic forcings. *Environmental Research Letters*, *12*(2), 024009.

- Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated.

- Köppen, W. P., & Geiger, R. (1923). Klimakarte der erde. Justus Perthes.

- Konda, G., & Vissa, N. K. (2023). Evaluation of CMIP6 models for simulations of surplus/deficit summer monsoon conditions over India. *Climate Dynamics*, 60(3-4), 1023-1042.

- Krishnan, R., Sanjay, J., Gnanaseelan, C., Mujumdar, M., Kulkarni, A., & Chakraborty, S. (2020). *Assessment of climate change over the Indian region: a report of the ministry of earth sciences (MOES), government of India* (p. 226). Springer Nature.

- Krishnan, R., Sabin, T. P., Vellore, R., Mujumdar, M., Sanjay, J. et al. (2016) Deciphering the desiccation trend of the South Asian monsoon hydroclimate in a warming world. *Climate Dynamics*, 47 (3), pp.1007-1027. ff10.1007/s00382-015-2886-5ff. ffhal-01322856f.

- Kratz, M., & Resnick, S. I. (1996). The QQ-estimator and heavy tails. *Stochastic Models*, 12(4), 699-724.

- Kulkarni, M. A., Singh, A., & Mohanty, U. C. (2012). Effect of spatial correlation on regional trends in rain events over India. *Theoretical and Applied Climatology*, *109*, 497-505.

- Langousis, A., Mamalakis, A., Puliga, M., & Deidda, R. (2016). Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research*, *52*(4), 2659-2681. https://doi.org/10.1002/2015WR018502.

- Lai, W., Wang, H., & Zhang, J. (2018). Comprehensive assessment of drought from 1960 to 2013 in China based on different perspectives. *Theoretical and Applied Climatology*, 134, 585-594.

- Lehmann, E.L. (1997). *Testing Statistical Hypotheses.* 2nd edition, Springer, New York

- Li, C., Singh, V. P., & Mishra, A. K. (2012). Simulation of the entire range of daily precipitation using a hybrid probability distribution. *Water resources research*, *48*(3). W03521. doi:10.1029/2011WR011446.

- Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American statistical association*, *9*(70), 209-219.

- Malamud, B. D. (2004). Tails of natural hazards. *Physics World*, *17*(8), 25.

- Manz, K., & Mansmann, U. (2020). Distributional challenges regarding data on death and incidences during the SARS-CoV-2 pandemic up to July 2020. *medRxiv*, 2020-07.

- May, W., Joseph, K. K., & Nkomoki, J. The variability and extremes of daily precipitation at 38 meteorological stations operated by the. *Danish Climate Centre Report 13-03*.

- Martel, B., El Adlouni, S., & Bobée, B. (2013). Comparison of the power of lognormality tests with different right-tail alternative distributions. *Journal of Hydrologic Engineering*, *18*(1), 1-9.

- Madsen, H., Rasmussen, P. F., & Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. *Water resources research*, *33*(4), 747-757.

- Mailhot, A., Lachance-Cloutier, S., Talbot, G., & Favre, A. C. (2013). Regional estimates of intense rainfall based on the Peak-Over-Threshold (POT) approach. *Journal of Hydrology*, *476*, 188-199. https://doi.org/10.1016/j.jhydrol.2012.10.036.

- Makkonen, L. (2006). Plotting positions in extreme value analysis. *Journal of Applied Meteorology and Climatology*, *45*(2), 334-340. https://doi.org/10.1175/JAM2349.1.

- Masaki, Y., Hanasaki, N., Takahashi, K., & Hijioka, Y. (2014). Global-scale analysis on future changes in flow regimes using Gini and Lorenz asymmetry coefficients. *Water Resources Research*, *50*(5), 4054-4078.

- Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association, 46*(253), 68–78.

- Maurer, E. P., & Hidalgo, H. G. (2008). Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrology and Earth System Sciences*, *12*(2), 551-563. 10.5194/hessd-4-3413-2007.

- Marsaglia, G., Tsang, W. & Wang, J. (2003). Evaluating Kolmogorov's Distribution. *Journal of Statistical Software,* 8(18).

- Martha, T. R., Roy, P., Govindharaj, K. B., Kumar, K. V., Diwakar, P. G., & Dadhwal, V. K. (2015). Landslides triggered by the June 2013 extreme rainfall event in parts of Uttarakhand state, India. *Landslides*, 12, 135-146.

- Mayooran, T., Laheetharan, A., (2014). The statistical distribution of annual maximum rainfall in Colombo district. *Sri Lankan Journal of Applied Statistics*, *15*(2), 107-130.

- McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.

- Methni, J. E., Gardes, L., & Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, *41*(4), 988-1012.

- Meehl, G. A., Hu, A. & Santer, B. D. (2008). The Mid-1970s Climate Shift in the Pacific and the Relative Roles of Forced versus Inherent Decadal Variability. *Journal of Climate,* 22, 780–792. https://doi.org/10.1175/2008JCLI2552.1

- Michalewicz, Z., Janikow, C. Z., & Krawczyk, J. B. (1992). A modified genetic algorithm for optimal control problems. *Computers and mathematics with applications,* 23(12), 83-94.

- Mielke Jr, P. W. (1973). Another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology and Climatology*, *12*(2), 275-280.

- Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, *22*(1), 67.

- Mishra, V., Kumar, D., Ganguly, A. R., Sanjay, J., Mujumdar, M., Krishnan, R., & Shah, R. D. (2014). Reliability of regional and global climate models to simulate precipitation extremes over India. *Journal of Geophysical Research: Atmospheres*, *119*(15), 9301-9323.

- Miller, A. J., Cayan, D. R., Barnett, T. P., Graham, N. E., & Oberhuber, J. M. (1994). The 1976-77 climate shift of the Pacific Ocean. *Oceanography*, 7(1), 21-26.

- Mishra, S. K., Sahany, S., & Salunke, P. (2017). Linkages between MJO and summer monsoon rainfall over India and surrounding region. *Meteorology and Atmospheric Physics*, *129*, 283-296.

- Mishra, V., Bhatia, U., & Tiwari, A. D. (2020). Bias-corrected climate projections for South Asia from coupled model intercomparison project-6. *Scientific data*, *7*(1), 338.

- Mlyński, D., Wałęga, A., Petroselli, A., Tauro, F. & Cebulska, M. (2019). Estimating maximum daily precipitation in the Upper Vistula Basin, Poland. *Atmosphere (Basel)* 10, 43. https://doi.org/10.3390/atmos10020043.

- Moccia, B., Mineo, C., Ridolfi, E., Russo, F., & Napolitano, F. (2021). Probability distributions of daily rainfall extremes in Lazio and Sicily, Italy, and design rainfall inferences. *Journal of Hydrology: Regional Studies*, *33*, 100771.

- Mondal, A., & Mujumdar, P. P. (2015). Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India. *Journal of Hydrology*, *521*, 217-231.

- Monjo, R., & Martin-Vide, J. (2016). Daily precipitation concentration around the world according to several indices. *International Journal of Climatology*, *36*(11), 3828-3838.

- Mukherjee, S., Aadhar, S., Stone, D., & Mishra, V. (2018). Increase in extreme precipitation events under anthropogenic warming in India. *Weather and climate extremes*, *20*, 45-53.

- Nagaev, A., & Tsitsiashvili, G. (2006). Tail asymptotics of the nth convolution of super-exponential distributions. *Statistics & probability letters*, *76*(9), 861-870.

- Nair, N. U., Nair, K. M., & Sreelakshmi, N. (2012). *Some properties of the new Zenga curve* (p. 43). Vita e pensiero.

- Nash, D. B. (1994). Effective sediment-transporting discharge from magnitude-frequency analysis. *The Journal of Geology*, *102*(1), 79-95.

- Nassa, R. A. K., Kouassi, A. M., & Toure, M. L. (2021). Sensitivity of Statistical Models for Extremes Rainfall Adjustment Regarding Data Size: Case of Ivory Coast. *Journal of Water Resource and Protection*, *13*(8), 654-674.

- Nandargi, S., Gaur, A., & Mulye, S. S. (2016). Hydrological analysis of extreme rainfall events and severe rainstorms over Uttarakhand, India. *Hydrological Sciences Journal*, *61*(12), 2145-2163. DOI: 10.1080/02626667.2015.1085990

- Németh, L., & Zempléni, A. (2020). Regression estimator for the tail index. *Journal of Statistical Theory and Practice*, *14*(3), 48. https://doi.org/10.1007/s42519-020-00114-7.

- Nerantzaki, S. D., & Papalexiou, S. M. (2019). Tails of extremes: Advancing a graphical method and harnessing big data to assess precipitation extremes. *Advances in Water Resources*, *134*, 103448. doi: https://doi.org/10.1016/j.advwatres.2019.103448.

- Nerantzaki, S., & Papalexiou, S. M. (2021). Assessing Extremes in Hydroclimatology: A Review on Probabilistic Methods, *Journal of Hydrology*, doi: https://doi.org/10.1016/j.jhydrol.2021.127302

- Nerantzaki, S. D., & Papalexiou, S. M. (2022). Assessing extremes in hydroclimatology: A review on probabilistic methods. Journal of Hydrology, 605, 127302.

- Nikulin, G., Kjellstro¨ M, E., Hansson, U. L. F., Strandberg, G., Ullerstig, A. (2011). Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. *Tellus A: Dynamic Meteorology and Oceanography*, 63(1), 41-55.

- Nieboer, D. (2011). Heuristics of heavy-tailed distributions and the Obesity index. *Dissertation*. Delft University of Technology.

- Nordhaus, W., Azam, Q., Corderi, D., Hood, K., Victor, N. M., Mohammed, M., Miltner, A., & Weiss, J. (2006). The g-econ database on gridded output: Methods and data. *Yale university*.

- O'Kane, T. J., Matear, R. J., Chamberlain, M. A., Oke, P. R. (2014). ENSO regimes and the late 1970's climate shift: The role of synoptic weather and South Pacific ocean spiciness. *Journal of Computational Physics,* 271, 19-38.

- O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... & Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461-3482.

- Oruc, S. (2022). Performance of bias corrected monthly CMIP6 climate projections with different reference period data in Turkey. *ActaGeophysica*, 70(2), 777-789.

- Ouarda, T.B.M.J., Ashkar, F., Bensaid, E., & Hourani, I. (1994) Statistical distributions used in hydrology. Transformations and asymptotic properties. *Scientific Report*, 31 pp., Department of Mathematics, Univ. of Moncton, New Brunswick.

- Ozonur, D., Pobocikova, I., & de Souza, A. (2021). Statistical analysis of monthly rainfall in Central West Brazil using probability distributions. *Modeling Earth Systems and Environment*, *7*, 1979-1989.

- Pai, D., Sridhar, L., Rajeevan, M., Sreejith, O., Satbhai, N., & Mukhopadhyay, B. (2014). Development of a new high spatial resolution (0.25 × 0.25) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam*, 65: 1–18. http://www.imd.gov.in/advertisements/ 20170320_advt_34.

- Panahi, H. (2016). Model selection test for the heavy-tailed distributions under censored samples with application in financial data. *International Journal of Financial Studies*, ISSN 2227-7072, M.D.P.I., Basel, Vol. 4, Iss. 4, pp 1-14. doi:10.3390/ijfs4040024.

- Panorska, A.K., Gershunov, A., & Kozubowski, T.J. (2007). From diversity to volatility: probability of daily precipitation extremes. *Nonlinear dynamics in geophysics*. Springer, New York, pp 465–484.

- Papalexiou, S.M., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, *17*(2), 851–862. doi: org/10.5194/hess-17-851-2013.

- Papalexiou, S.M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, *49*, 187–201. doi:10.1029/2012WR012557.

- Papalexiou, S.M., & Koutsoyiannis, D. (2016). A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Advances in Water Resources*, *94*, 131–145. https://doi.org/10.1016/J.ADVWATRES.2016.05.005.

- Papalexiou, S.M., AghaKouchak, A., & Foufoula-Georgiou, E. (2018). A Diagnostic Framework for Understanding Climatology of Tails of Hourly Precipitation Extremes in the United States. *Water Resources Research*, *54*, 6725–6738. https://doi.org/10.1029/2018WR022732.

- Papalexiou, S. M., & Montanari, A. (2019). Global and regional increase of precipitation extremes under global warming. *Water Resources Research*, *55*(6), 4901-4914.

- Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences*, *11*(5), 1633-1644.

- Phien, H. N., Ajirajah, T. J., (1984). Applications of the log Pearson type-3 distribution in hydrology. *Journal of hydrology*, *73*(3-4), 359-372.

- Pickands, J. I. (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*3, 119-131.

- Piani, C., Haerter, J. O., & Coppola, E. (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and applied climatology*, *99*, 187-192.

- Preisendanz, H. E., Veith, T. L., Zhang, Q., Shortle, J., (2020). Temporal inequality of nutrient and sediment transport: a decision-making framework for temporal targeting of load reduction goals. *Environmental Research Letters*, *16*(1), 014005.

- Prajeesh, A. G., Swapna, P., Krishnan, R., Ayantika, D. C., Sandeep, N., Manmeet, S., ...& Sandip, I. (2021). The Indian summer monsoon and Indian Ocean dipole connection in the IITM Earth system model (IITM-ESM). *Climate Dynamics*, 1-21.

- Qi, Y. (2008). Bootstrap and empirical likelihood methods in extremes. *Extremes, 11*, 81–97. https://doi.org/10.1007/s10687-007-0049-8.

- Rajeevan, M., Bhate, J., & Jaswal, A.K. (2008). Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophysical Research Letters*, 35, L18707.

- Rajbanshi, J., & Das, S. (2021). The variability and teleconnections of meteorological drought in the Indian summer monsoon season: Implications for staple crop production. *Journal of Hydrology*, 603, 126845.

- Rajah, K., O'Leary, T., Turner, A., Petrakis, G., Leonard, M., & Westra, S. (2014). Changes to the temporal distribution of daily precipitation. *Geophysical Research Letters*, *41*(24), 8887-8894.

- Rajulapati, C. R., Papalexiou, S. M., Clark, M. P., Razavi, S., Tang, G., & Pomeroy, J. W. (2020). Assessment of extremes in global precipitation products: How reliable are they?.*Journal of Hydrometeorology*, *21*(12), 2855-2873.

- Rajulapati, C. R., & Papalexiou, S. M. (2023). Precipitation Bias Correction: A Novel Semi-parametric Quantile Mapping Method. *Earth and Space Science*, *10*(4), e2023EA002823.https://doi.org/10.1029/2023EA002823.

- Rai, P., Choudhary, A., & Dimri, A. P. (2019). Future precipitation extremes over India from the CORDEX-South Asia experiments. *Theoretical and Applied Climatology*, *137*(3), 2961-2975.

- Rai, P. K., Singh, G. P., & Dash, S. K. (2020). Projected changes in extreme precipitation events over various subdivisions of India using RegCM4. *Climate Dynamics*, *54*(1), 247-272.

- Rao, K.K., Patwardhan, S.K., Kulkarni, A., Kamala, K., Sabade, S.S., & Kumar, K.K. (2014). Projected changes in mean and extreme precipitation indices over India using PRECIS. *Global and Planetory Change*, 113, 77–90.

- Rao, K. K., Kulkarni, A., Patwardhan, S., Kumar, B. V., & Kumar, T. L. (2020). Future changes in precipitation extremes during northeast monsoon over south peninsular India. *Theoretical and Applied Climatology*, *142*(1), 205-217.

- Reddy, N. M., & Saravanan, S. (2023). Extreme precipitation indices over India using CMIP6: a special emphasis on the SSP585 scenario. *Environmental Science and Pollution Research*, *30*(16), 47119-47143.

- Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

- Reeve, D. E. (1996). Estimation of extreme Indian monsoon rainfall. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *16*(1), 105-112.

- Roxy, M.K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., Terray, P., & Rajeevan, M. (2017). A threefold rise in widespread extreme rain events over central India. *Nature Communication*, *8*, 1–11. https://doi.org/ 10.1038/s41467-017-00744-9.

- Rosenzweig, C., Iglesius, A., Yang, X. B., Epstein, P. R., & Chivian, E. (2001). *Climate change and extreme weather events-Implications for food production, plant diseases, and pests*.NASA Publications, 24.

- Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E., & Bloomfield, J. (2002). Increased crop damage in the US from excess precipitation under climate change. *Global Environmental Change*, *12*(3), 197–202. https://doi.org/10.1016/S0959-3780(02)00008-0.

- Roth, M., Jongbloed, G., & Buishand, T.A. (2016). Threshold selection for regional peaks-over threshold data. *Journal of Applied Statistics*, *43*, 1291–1309. https://doi.org/10.1080/02664763.2015.1100589.

- Rolski, T., Schmidli, H., Schmidt, V., & Teugels, J. L. (2009). *Stochastic processes for insurance and finance*. John Wiley & Sons.

- Rubel, F., & Kottek, M. (2010). Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *MeteorologischeZeitschrift*, *19*(2), 135.

- Sabeerali, C. T., Rao, S. A., Ajayamohan, R. S., & Murtugudde, R. (2012). On the relationship between Indian summer monsoon withdrawal and Indo- Pacific SST anomalies before and after 1976/1977 climate shift. *Climate Dynamics, 39*(3), 841–859. https://doi.org/10.1007/s00382-011-1269-9.

- Sabeerali, C. T., Ajayamohan, R. S., Bangalath, H. K., & Chen, N. (2019). Atlantic zonal mode: an emerging source of Indian summer monsoon variability in a warming world. *Geophysical Research Letters*, *46*(8), 4460-4467.

- Sahana, A. S., Ghosh, S., Ganguly, A., & Murtugudde, R. (2015). Shift in Indian summer monsoon onset during Shift in Indian summer monsoon onset during 1976/1977. *Environmental Research Letters, 10*(5), 054006. https:// doi.org/10.1088/1748-9326/10/5/054006.

- Sarkar, S. & Maity, R. (2020) Increase in Probable Maximum Precipitation in a Changing Climate over India. *Journal of Hydrology.* doi: https://doi.org/10.1016/j.jhydrol.2020.124806.

- Sarkar, S., & Maity, R. (2021). Global climate shift in 1970s causes a significant worldwide increase in precipitation extremes. *Scientific reports*, *11*(1), 1-11.

- Sarkar, S., & Maity, R. (2022). Future Characteristics of Extreme Precipitation Indicate the Dominance of Frequency Over Intensity: A Multimodel Assessment From CMIP6 Across India. *Journal of Geophysical Research: Atmospheres, 127*(16), e2021JD035539.

- Saha, A., Ghosh, S., Sahana, A. S., & Rao, E. P. (2014). Failure of CMIP5 climate models in simulating post-1950 decreasing trend of Indian monsoon. *Geophysical Research Letters*, *41*(20), 7323-7330.

- Saha, U., & Sateesh, M. (2022). Rainfall extremes on the rise: Observations during 1951–2020 and bias-corrected CMIP6 projections for near-and late 21st century over Indian landmass. *Journal of Hydrology*, *608*, 127682.

- Sartori, M., & Schiavo, S., (2015). Connected we stand: a network perspective on trade and global food security. *Food Policy*, *57*, 114–127. https://doi.org/10.1016/j.foodpol.2015.10.004.

- Sannan, M. C., Nageswararao, M. M., & Mohanty, U. C. (2020). Performance evaluation of CORDEX-South Asia simulations and future projections of northeast monsoon rainfall over south peninsular India. *Meteorology and Atmospheric Physics*, 1-28.

- Sanjay, J., Ramarao, M. V. S., Mujumdar, M., & Krishnan, R. (2017). Regional climate change scenarios. *Observed climate variability and change over the Indian region*, 285-304. doi:10.1007/978-981-10-2531-0.

- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

- Serinaldi, F. & Kilsby, C.G. (2014). Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resources Research*, 50, 336–352. https://doi.org/10.1002/2013WR014211.

- Shastri, H., Paul, S., Ghosh, S. & Karmakar S. (2015). Impacts of urbanization on Indian summer monsoon rainfall extremes. *Journal of Geophysical Research: Atmospheres* 120,495–516, doi:10.1002/2014JD022061.

- Sharma, P. J., Patel, P. L., & Jothiprakash, V. (2020). Hydroclimatic teleconnections of large-scale oceanic-atmospheric circulations on hydrometeorological extremes of Tapi Basin, India. *Atmospheric Research*, 235, 104791.

- Sharma, M. A., & Singh, J. B. (2010). Use of probability distribution in rainfall analysis. New York Science Journal, 3(9), 40-49.

- Sharma, M., Coulibaly, P., & Dibike, Y. (2011). Assessing the need for downscaling RCM Data for Hydrologic Impact Study. *Journal of Hydrologic Engineering*, *16*, 534-539. 10.1061/(ASCE)HE.1943-5584.0000349.

- Sharmila, S., Joseph, S., Sahai, A.K., Abhilash, S., & Chattopadhyay, R. (2015). Future projection of Indian summer monsoon variability under climate change scenario: An assessment from CMIP5 climate models. *Global and Planetary Change*, *124*, 62-78, ISSN 0921-8181. https://doi.org/10.1016/j.gloplacha.2014.11.004.

- Shastri, H., Paul, S., Ghosh, S., & Karmakar S. (2015). Impacts of urbanization on Indian summer monsoon rainfall extremes. *Journal of Geophysical Research: Atmospheres,* 120,495–516, doi:10.1002/2014JD022061.

- Shrestha, S., Yao, T., Kattel, D. B., & Devkota, L. P. (2019). Precipitation characteristics of two complex mountain river basins on the southern slopes of the central Himalayas. *Theoretical and Applied Climatology*, *138*(1), 1159-1178.

- Sherif, M., Almulla, M., Shetty, A., & Chowdhury, R. K. (2014). Analysis of rainfall, PMP and drought in the United Arab Emirates. *International journal of climatology*, *34*(4), 1318-1328.

- Smith, J. A., Cox, A. A., Baeck, M. L., Yang, L., & Bates, P., 2018. Strange floods: The upper tail of flood peaks in the United States. Water Resources Research, 54(9), 6510-6542.

- Stanley, M. H., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A., & Stanley, H. E. (1995). Zipf plots and the size distribution of firms. *Economics letters*, *49*(4), 453-457.

- Strupczewski, W. G., Kochanek, K., Markiewicz, I., Bogdanowicz, E., Weglarczyk, S., & Singh, V. P., 2011. On the tails of distributions of annual peak flow. Hydrology Research, 42(2-3), 171-192.

- Singh, J., Vittal, H., Karmakar, S., Ghosh, S., & Niyogi, D. (2016) Urbanization causes non-stationarity in Indian summer monsoon rainfall extremes. *Geophysical Research Letters, 43*(21), 11-269.

- Singh, D., Tsiang, M., Rajaratnam, B., & Diffenbaugh, N.S. (2014) Observed changes in extreme wet and dry spells during the south Asian summer monsoon season. *Nature Climate Change*, *4*, 456–461. https://doi.org/10.1038/nclimate2208.

- Suman, M., & Maity, R. (2020). Southward shift of precipitation extremes over south Asia: Evidences from CORDEX data. *Scientific reports*, *10*(1), 1-11.

- Supharatid, S., Nafung, J., & Aribarg, T. (2021). Projected changes in temperature and precipitation over mainland Southeast Asia by CMIP6 models. *Journal of Water and Climate Change*, 13, 1–20. https://doi.org/10.2166/wcc.2021.015.

- Suman, M., Maity, R., & Kunstmann, H. (2022). Precipitation of Mainland India: Copula-based bias-corrected daily CORDEX climate data for both mean and extreme values.

- Taylor, K. E., Stouffer, R.J., & Meehl, G.A. (2012). An overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*, 485-498. https://doi.org/10.1175/BAMS-D-11-00094.1.

- Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific reports*, *10*(1), 1-10.

- Teugels, J.L. (1975). The Class of Subexponential Distributions. *The Annals of Probability*. doi:10.2307/2959204.

- Theriault, V., & Serra, R. (2014). Institutional environment and technical efficiency: A stochastic frontier analysis of cotton producers in West Africa. *Journal of Agricultural Economics*, *65*(2), 383-405.

- Tong, Q., Swallow, B., Zhang, L., & Zhang, J. (2019). The roles of risk aversion and climate-smart agriculture in climate risk management: Evidence from rice production in the Jianghan Plain, China. *Climate Risk Management*, *26*, 100199.

- Todmal, R.S. (2021). Future Climate Change Scenario over Maharashtra, Western India: Implications of the Regional Climate Model (REMO-2009) for the Understanding of Agricultural Vulnerability. *Pure and Applied Geophysics*, *178*, 155–168. https://doi.org/10.1007/s00024-020-02642-6.

- Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate research*, *47*(1-2), 123-138.

- Turco, M., Antonella, S., Herrera, G., Sixto, Maria, L., & Gutiérrez, J. (2013). Large biases and inconsistent climate change signals in ENSEMBLES regional projections. *Climatic Change, 120*. 859-869. 10.1007/s10584-013-0844-y.

- Varikoden, H., Revadekar, J. V., Kuttippurath, J., & Babu, C. A. (2019). Contrasting trends in southwest monsoon rainfall over the Western Ghats region of India. *Climate Dynamics*, *52*(7), 4557-4566.

- Vargo, E., Pasupathy, R., & Leemis, L. M., (2010). Moment-ratio diagrams for univariate distributions. *Journal of Quality Technology*, *42*, 276-286.

- Vela, A.C. & Rodríguez, G. (2014). *Extreme Value Theory: An Application to the Peruvian Stock Market Returns*. Documentos de Trabajo / Working Papers.

- Villarini, G., Smith, J. A., Ntelekos, A. A. & Schwarz U. (2011a). Annual maximum and peaks-over-threshold analyses of daily rainfall accumulations for Austria. *Journal of Geophysical Research,* 116, D05103. doi:10.1029/2010JD015038.

- Villarini, G., Smith, J. A., Baeck, M. L., Marchock, T., & Vecchi, G. A. (2011b). Characterization of rainfall distribution and flooding associated with U.S. landfalling tropical cyclones: Analyses of Hurricanes Frances, Ivan, and Jeanne (2004). *Journal of Geophysical Research*, 116, D23116. https://doi.org/10.1029/2011JD016175.

- Villarini, G. (2012). Analyses of annual and seasonal maximum daily rainfall accumulations for Ukraine, Moldova, and Romania. *International Journal of Climatology*, *32*(14), 2213-2226.

- Vittal, H., Karmakar, S., & Ghosh, S. (2013). Diametric changes in trends and patterns of extreme rainfall over India from pre-1950 to post-1950. *Geophysical Research Letters* , *40*(12), 3253-3258.

- Vinnarasi, R., & Dhanya, C. T. (2016). Changing characteristics of extreme wet and dry spells of Indian monsoon rainfall. *Journal of Geophysical Research: Atmospheres*, *121*(5), 2146-2160. doi:10.1002/2015JD024310

- Vinod, D., & Agilan, V. (2022). Impact of Climate Change on Precipitation Over India Using CMIP-6 Climate Models. In *Innovative Trends in Hydrological and Environmental Systems: Select Proceedings of ITHES 2021* (pp. 155-164). Singapore: Springer Nature Singapore.

- Villarini, G., & Smith, J.A., (2010). Flood peak distributions for the eastern United States. Water Resources Research, 46 (6), 1–17. doi:10.1029/2009WR008395.

- Vishnu Priya, M. S., & Agilan, V. (2022). Evaluation of change factor methods in downscaling extreme precipitation over India. *Journal of Hydrology*, *614*, 128531.

- Voitalov, I., Van Der Hoorn, P., Van Der Hofstad, R., & Krioukov, D. (2019). Scale-free networks well done. *Physical Review Research*, *1*(3), 033034. https://arxiv.org/abs/1811.02071.

- Wang, B., Wu, R. & Lau, K. M. (2001). Interannual Variability of the Asian Summer Monsoon: Contrasts between the Indian and the Western North Pacific – East Asian Monsoons. *Journal of Climate, 14*, 4073–4090.

- Wang, H. (2001). The weakening of the Asian monsoon circulation after the end of 1970's. *Advances in Atmospheric Sciences, 18*(3), 376–386. https://doi.org/10.1007/BF02919316.

- Wang, Y., Xu, H., Cheng, D. & Yu, C. (2018). The local asymptotic estimation for the supremum of a random walk with generalized strong subexponential summands. *Statistical Papers, 59*(1), 99-126.

- Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., ... & Roberts, N. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, *52*(3), 522-555.

- Weibull, W. (1939). A statistical theory of strength of materials. Ing. Vetensk. Akad. Handl*, 151*, 1–45.

- Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A. & Palmer, T. (2017). Atmospheric seasonal forecasts of the twentieth century: Multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Quarterly Journal of Royal Meteorological Society, 143*, 917–926. https://doi.org/10.1002/qj.2976.

- Werner, T., & Upper, C. (2004). Time variation in the tail behavior of Bund future returns. *Journal of Futures Markets, 24*(4): 387–398. doi:10.1002/fut.10120.

- Wilson, P. S., & Toumi, R., (2005). A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters*, 32(14).

- Wietzke, L. M. , Merz , B., Gerlitz , L., Kreibich , H., Guse , B., Castellarin A. & Vorogushyn S. (2020) Comparative analysis of scalar upper tail indicators. *Hydrological Sciences Journal*, *65*(10): 1625-1639. doi: 10.1080/02626667.2020.1769104.

- Xie, X. (2017). Analysis of Heavy-Tailed Time Series (Doctoral dissertation, University of Copenhagen, Faculty of Science, Department of Mathematical Sciences).

- Yaduvanshi, A., Bendapudi, R., Nkemelang, T., & New, M. (2021). Temperature and rainfall extremes change under current and future warming global warming levels across Indian climate zones. *Weather and Climate Extremes*, *31*, 100291. https://doi. org/10.1016/j.wace.2020.100291.

- Zarrin, A., & Dadashi-Roudbari, A. (2021). Projection of future extreme precipitation in Iran based on CMIP6 multi-model ensemble. *Theoretical and Applied Climatology*, *144*, 643-660.

- Zenga, M., (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, *5*(1), 3-27.

- Zhang, Q., Gu, X., Singh, V. P., Xu, C. Y., Kong, D., Xiao, M., & Chen, X. (2015). Homogenization of precipitation and flow regimes across China: changing properties, causes and implications. *Journal of Hydrology*, *530*, 462–475. doi:10.1016/j.jhydrol.2015.09.041.

- Zhang, R. H., Rothstein, L. M. & Busalacchi, A. J. (1998). Origin of upper ocean warming and El Niño change on decadal scales in the tropical Pacific Ocean. *Nature 391*, 879–83.

- Zhou, T., Gong, D., Li, J. & Li, B. (2009). Detecting and understanding the multi-decadal variability of the East Asian Summer Monsoon–Recent progress and state of affairs. *Meteorologische Zeitschrift* , *18*(4), 455-467.

- Zhanling, Li., Zhanjie, Li., Zhao, W., & Wang, Y. (2015). Probability Modeling of Precipitation Extremes over Two River Basins in Northwest of China. *Advances in Meteorology*. Article ID 3374127. http://dx.doi.org/10.1155/2015/374127.

- Zuo, J., Ren, H.-L., Li, W., & Wang, L. (2016). Interdecadal variations in the relationship between the winter North Atlantic oscillation and temperature in south-central China. *Journal of Climate, 29*, 7477–7493. https://doi.org/10.1175/JCLI-D-15-0873.1.

APPENDIXES

Appendix A: Papalexiou et al. (2013) and Nerantzaki and Papalexiou (2019) proposed threshold-based approaches to characterize the tail behavior of daily precipitation. A brief description of the threshold-based approach is presented below

A.1 Threshold-based approach proposed by Papalexiou et al. (2013)

In this approach, the annual exceedance series (AES) (e.g., Chow 1964; Ben-Zvi, 2009) comprising N extreme precipitation events above some threshold x_L is extracted from the non-zero daily precipitation data recorded at a station, where N equals the number of recording years. Then, the empirical probability of exceedance $\bar{F}_N(x_i)$ is calculated using the Weibull plotting position formula (Weibull, 1939; Makkonen, 2006) as given by Equation A.1.

$$\bar{F}_N(x_i) = 1 - \frac{r(x_i)}{n+1}$$
(A.1)

Where $r(x_i)$ is the rank of the precipitation value x_i in the ordered sample and n number of non-zero precipitation values. The empirical tail of daily precipitation data is represented by the N largest non-zero precipitation values (i.e., AES series) and denoted by $\bar{F}_N(x_i)$ from $n - N + 1 \leq i \leq n$ (note that $x_L = x_{(n-N+1)}$). Subsequently, the theoretical tail represented by the probability of exceedance $\bar{F}(x_i)$ is determined for each of the four representative distributions given in Table 3.1 of Chapter 3.1. Finally, a PRMSE norm given by Equation A.2 is used to evaluate the fit and compare the performance of different theoretical tails to the empirical tails.

$$PRMSE = \frac{1}{N} \sum_{i=n-N+1}^{n} \left(\frac{\bar{F}(x_{(i)})}{\bar{F}_N(x_{(i)})} - 1 \right)^2$$
(A.2)

The distribution yielding the least PRMSE is chosen as the best-suited probability distribution for the tail of precipitation data (Papalexiou et al., 2013; Moccia et al., 2021). A detailed description of the approach can be found in Papalexiou et al. (2013).

A.2 Threshold-based approach proposed by Nerantzaki and Papalexiou (2019)

Nerantzaki and Papalexiou (2019) proposed the advancement of a popular graphical method known as the mean excess function (MEF) to discriminate the tails of daily precipitation at 21,348 stations from all over the globe. For a random variable X with a distribution function, F_X the mean excess function $e(x_p)$ is given by

$$e(x_p) = E(X - x_p \mid X > x_p) = \frac{1}{\overline{F}_X(x_p)} \int_{x_p}^{\infty} (x - x_p) f_X(x)\, dx, \, x_p \geq 0 \tag{A.3}$$

where, $x_p = F_X^{-1}(p)$ is the lower threshold value corresponding to probability p and $F_X^{-1}(p)$ is the quantile function.

In this approach, empirical $e(x_p)$ is estimated for various threshold values $x_p \in (x_{0.9}, x_{0.99})$ at each station. The non-exceedance probabilities from 0.9 to 0.99 were considered as the tail of precipitation probability. The plot between x_p and $e(x_p)$ are then constructed, and the linear slope of the plot is determined. The slope of the plot forms the basis to determine whether the tail is exponential (i.e., slope close to zero) or not. Nerantzaki and Papalexiou (2019) considered two statistical tests in their study that are followed in this study also. Firstly, a two-tailed test is performed to test the null hypothesis that the tail of daily precipitation at a station is exponential. The test utilizes a confidence interval (CI) for the exponential case corresponding to sample size (n) for a specific significance level (u). If the estimated slope lies within the CI, then the null hypothesis cannot be rejected for the selected significance level. Contrary to this, if the estimated slope lies outside the CI, then the null hypothesis is rejected, indicating a non-exponential tail. When the null hypothesis from the two-tailed test is rejected, a one-tailed test is performed where the null hypothesis is Exponential tail, and the alternative is sub-exponential (or hyper-exponential). Tails heavier than the exponential tail, i.e., the observed slope is above the CI's upper limit are designated as the sub-exponential tails, whereas the tails lighter than the exponential tail, i.e., the observed slope is below the CI's upper limit are designated as the hyper-exponential tails. Finer details about the approach can be found in Nerantzaki and Papalexiou (2019).