Analysis of Extreme Events in the Environment

A Project Report Submitted in Partial Fulfillment of the

Requirement for the fourth semester **M.Sc. (Statistics)**

('A' Grade NAAC Re-accredited fourth Cycle)

submitted by

Ms. Devyani Sunil Bhosle (Seat No.358777)

Ms. Mayuri Sanjay Patil (Seat No.358790)

Ms. Harshada Jagdish Shirsale (Seat No.358798)

Under the Guidance of

Prof. Kirtee K. Kamalja

Department of Statistics

School of Mathematical Sciences

Kavayitri Bahinabai Chaubhari

North Maharashtra University, Jalgaon

(Academic Year: 2024 - 25)

<u>CERTIFICATE</u>

This is to certify that **Ms. Devyani Sunil Bhosle**, **Ms. Mayuri Sanjay Patil**, **Ms. Harshada Jagdish Shirsale** Students of M.Sc.(Statistics) at Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon have successfully completed their project work entitled **"Analysis of Extreme Events in Environment"**. This project was completed as part of M.Sc. (Statistics) program under my guidence and supervision during the academic year December 2024 to May 2025.

Prof. Kirtee K. Kamalja
(Project Guide)

ABSTRACT

The study entitled “Analysis of Extreme Events in Environment” investigates the occurrence, characteristics, and impacts of extreme environmental events using advanced statistical methods, particularly those from the domain of Extreme Value Theory (EVT). These events—ranging from floods, droughts, and heatwaves to cyclones, wildfires, and cold waves—pose significant threats to human life, infrastructure, ecosystems, and national economies. The project primarily aims to understand the severity, frequency, and socio-economic consequences of such events under the growing influence of climate change.

The report begins by exploring the concept of return periods, which indicate the average interval of time between extreme events of a given magnitude. Using return period calculations and rainfall data from various Indian regions, especially the high-rainfall state of Meghalaya, the study estimates the likelihood of future extreme rainfall occurrences. Statistical modeling is employed through Gumbel, Fréchet, and Weibull distributions (components of the Generalized Extreme Value Distribution), as well as the Generalized Pareto Distribution (GPD) for threshold exceedances. These distributions help estimate return levels, quantify risks, and understand the tail behavior of extreme data.

The study also examines climate-related mortality and death rates, using indicators such as the death rate per year and death rate per million, providing insights into population vulnerability. A historical review of major global and Indian disasters (e.g., 2004 Indian Ocean Tsunami, 2005 Mumbai floods, 2015–16 Indian droughts, 2018 Kerala floods, and recent European heatwaves) highlights the scale of human and economic losses. The report calculates event-specific mortality rates, demonstrating the disproportionate burden of extreme events on low-income and high-density regions.

Furthermore, the economic and health impacts of extreme weather are analyzed. These include direct costs (infrastructure damage, agricultural loss) and indirect effects (decreased productivity, increased healthcare burden, mental health issues, and displacement). Climate change is identified as a major multiplier of health risks, especially in vulnerable populations such as the elderly, poor, and those in coastal areas.

In the statistical section, the report utilizes Peaks Over Threshold (POT) methods and `R` packages (e.g., `evir`, `ismev`, `extRemes`) to fit EVT models, simulate rainfall extremes, and interpret return levels for decision-making. A case study of rainfall data from Meghalaya illustrates the application of GPD modeling and the derivation of return levels for extreme precipitation events.

The project concludes that statistical modeling of extreme events is essential for disaster preparedness, climate resilience, and public policy formulation. Accurate risk assessments

can help design early warning systems, climate-adaptive infrastructure, and targeted response plans. However, limitations such as data availability, reporting inconsistencies, and modeling uncertainty are acknowledged. The study recommends continuous data collection, regional calibration of models, and the integration of local knowledge for effective climate risk management.

ACKNOWLEDGEMENT

The completion of this project is a dream come true with the help of supporters. The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have got all this along with completion of our project. All that we have done is only due to best supervision for that we would really like to show gratitude towards them.

We would like to convey sincere regards to **Prof. Kirtee K. Kamalja**, Head of Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon for providing us an opportunity to perform project giving us all support and guidance. It was also a great pleasure to work under **Prof. Kirtee K. Kamalja** valuable guidance, kind suggestion, cooperation and constant encouragement, which entitled us to take every forward step in our project. It would be unfair to go without acknowledging our beloved friends at from the Department of Statistics K.B.C.N.M.U, Jalgaon for their precious support to us. Finally, yet importantly, we are thankful to our parents for their moral support and blessings.

Devyani Sunil Bhosle
Mayuri Sanjay Patil
Harshada Jagdish Shirsale

Contents

List of Figures

List of Tables

Chapter 1:
About Extreme Events

Pinheiro and Grotjahn [2015]Extreme events, whether natural or human-induced, pose significant challenges to human societies, economies, and ecosystems. These events, characterized by their intensity, unpredictability, and widespread impact, have become increasingly frequent and severe due to factors such as climate change, urbanization, and environmental degradation. Natural extreme events include phenomena like floods, earthquakes, cyclones, droughts, and heatwaves, while human-induced events encompass industrial accidents, deforestation, and large-scale pollution incidents. Both types of events can lead to catastrophic consequences, including loss of life, economic disruptions, environmental degradation, and long-term societal impacts.

Understanding the nature, frequency, and severity of extreme events is essential for risk assessment, disaster preparedness, and the development of effective mitigation strategies. Statistical analysis provides a powerful toolset for this purpose, enabling researchers to identify patterns, measure impacts, and forecast future risks based on historical data. This study aims to explore extreme events from a statistical perspective, focusing on their classification, severity assessment, and socio-economic impacts. By leveraging statistical tools such as descriptive statistics and extreme value theory (EVT), the study seeks to offer data-driven insights that can inform policy-making, enhance disaster preparedness, and contribute to sustainable development goals.

The primary objectives of this project are to catalog various extreme events based on their nature (natural or human-induced) and regional occurrence, analyze historical data to identify trends, and develop a framework for assessing the severity of these events using indicators like casualties, economic losses, affected populations, and environmental damage. Furthermore, the study will select the most severe events for in-depth case studies, examining their social and economic consequences, focusing on long-term impacts and recovery processes. Visualizations such as charts and maps will be used to enhance data interpretation.

This research holds significant importance in the context of disaster preparedness, economic stability, social resilience, environmental sustainability, and climate change awareness. It aims to provide data-backed recommendations for disaster risk reduction, emergency response planning, and strategies to improve community resilience. However, the study acknowledges certain limitations, including data availability and quality, the complexity of quantifying long-term socio-economic impacts, and regional differences in data reporting standards.

The key research questions guiding this study include: What are the most common types of extreme events, and how frequently do they occur? Which events have the highest severity based on socio-economic and environmental impacts? How do extreme events influence economic growth, public health, and social stability? What statistical methods are most effective in predicting the occurrence and severity of extreme events? And finally, how can data-driven insights inform policy-making and disaster risk reduction strategies?

Through this comprehensive analysis, the study aims to contribute to a deeper understanding of extreme events, supporting efforts to build resilient communities, protect ecosystems, and foster sustainable development in the face of increasing environmental challenges.

1.1 The Concept of Return Period

You can learn about the return period, also known as the recurrence interval, from this source: Return Period - ScienceDirect. When defining the storm-induced hotspot, a relevant issue is the definition of its "severity". This is done through the selection of a return period for the analysis. More than one return period can be (and should be) computed for the same coastal area. This allows an evaluation of possible HS changes according to the return period used within each coastal area.

The chosen return periods will vary from site to site, depending on the return periods already in use for coastal management, and their selection should be agreed with local stakeholders. While in some countries (e.g., Portugal) return periods of 100 years are not yet (or rarely) considered, on highly protected coasts (e.g., Belgium) return periods greater than 1000 years are increasingly common in coastal management and safety plans.

The selection of return periods in the CRAF 1 should be discussed with stakeholders and reflect their needs or recommendations. The relatively limited number of years (few decades) of available measured or hindcast data reduces the ability to produce results with a high degree of accuracy for large return periods (hundreds to thousands of years), which is still a drawback of the CRAF methodology, as for any other.

On the other hand, this method permits results with a high degree of confidence for lower return periods (<100 years), which are most commonly used by the majority of coastal managers and end-users.

It may have been noted that both approaches to discretize a given time period into a number of time units or intervals are less user-friendly because there are terms (or variables) in their equations, such as p_{fi}, λ, p_{fe}, etc., that are difficult to obtain. This is in addition to some assumptions and approximations adopted in deriving these equations.

A practically natural approach is to discretize the time by year for stochastic processes. This approach is indeed widely employed in practice and is known as the return period (also referred to as recurrence interval or repeat interval).

There are many examples where the term return period is used to indicate the risk or severity of a risk for an event or hazard. The risk level of natural hazards—such as cyclones, earthquakes, and floods—is often measured by their return periods. Return period is a statistical measurement typically based on historical data over an extended period of time.

In general, a return period is the average time or an estimated average time between events, such as earthquakes, floods, and landslides. For example, the return period for design wind speed is usually 50 years in many design codes and standards. The historical return period for major hurricanes (i.e., wind speed > 49 m/s) is 58 years **?**. The most recent example is the COVID-19 pandemic, which is officially announced as a 1 in 100 years event, i.e. the return period is 100 years.

The basic idea of the return period approach is that, instead of considering actual time explicitly, the time between two failure events is of concern. Since the event is random, the time between two events is also random.

Let T be the time between two events, which is a random variable.The return period, denoted by $\bar{T}$ can be theoretically defined as follows:

The return period, denoted by $\bar{T}$ is defined as the expected time between two occurrences of a random event:

$$\bar{T} = \mathbb{E}[T] \tag{1.1}$$

The probability distribution of the random variable T can be established as follows. Consider a sequence of events. Since T is the time between two events (e.g., A and B), if one event (say, A) has occurred, the next event (B) should be the first occurrence in the subsequent sequence. Thus, T represents the time between these two events.

This scenario is analogous to a Bernoulli trial sequence, where each trial has two possible outcomes: occurrence or non-occurrence (success or failure in reliability terms). The time t at which the first occurrence (e.g., event B) takes place follows a geometric distribution (Devore, 1995):

$$P(T = t) = p(1 - p)^{t-1} \tag{1.2}$$

Here, p is the probability of occurrence of the event in any given time unit or trial. This model assumes that:

- Trials are independent.
- The outcome of one trial does not influence another.

Given the probability distribution, the expected return period (mean time between events) is calculated as:

$$\bar{T} = \mathbb{E}[T] = \sum_{t=1}^{\infty} t \cdot p(1 - p)^{t-1} = p\left[1 + \frac{2}{1 - p} + \frac{3}{(1 - p)^2} + \cdots\right] \tag{1.3}$$

From Taylor series expansion,

$$p\left(1 + \frac{2}{1 - p} + \frac{3}{(1 - p)^2} + \cdots\right) = \frac{1}{(1 - p)^2} - \frac{1}{1 - p} \tag{1.4}$$

Substituting this back gives:

$$\bar{T} = \mathbb{E}[T] = \frac{1}{p} \tag{1.5}$$

Equation (3.39) suggests that the return period between occurrences of events is the reciprocal of the probability of occurrence of an event in any given trial or time unit, i.e., p. This probability can be interpreted as the average frequency of the occurrences or the mean rate of occurrence.

It should be noted that the key assumptions in deriving the return period include:

1. The probability p is constant.
2. Events are independent.
3. Current events are independent of past events.

Although the term *return period* has been widely used, there still exists some misperception and misunderstanding of its meaning. For example, a 50-year earthquake does not mean that it will happen regularly every 50 years, or only once in 50 years. It only means that the probability of its occurrence in any one year is:

$$\frac{1}{50} = 0.02 \quad \text{or} \quad 2\% \text{ chance on average.}$$

The term "return period" may literally give a wrong connotation.Statistically, a 50-year event may occur once, twice, more than twice, or not at all within 50 years, and each occurrence is governed by probability.

More importantly, the return period is not the lifetime of a system or structure.These are two different concepts:

- Return period is the mean time between two occurrences of events a deterministic value.
- Lifetime is the time at which the probability of the event (e.g., structural failure) exceeds an acceptable level a probabilistic value.

1.2 Return period for the annual maximum rainfall

The concept of maximum rainfall refers to the greatest amount of rainfall recorded in a specific time period at a particular location. It is a key metric used in meteorology, hydrology, agriculture, and civil engineering for planning and risk assessment. Maximum rainfall is the highest amount of rain (measured in millimeters or inches) received in a particular: Minute, Hour, Day (24 hours), Month, Season, Year. It is usually measured using a rain gauge and reported as: mm/day,mm/hour.

- Maximum daily rainfall: Highest rainfall in 24 hours.

- Annual maximum rainfall: Highest daily rainfall each year, used in time series analysis.
- IDF analysis: Intensity-Duration-Frequency analysis uses maximum rainfall data to design drainage systems.
- Flood forecasting and disaster management
- Agricultural planning (e.g., sowing, irrigation)
- Infrastructure design (roads, bridges, dams)
- Climate studies

I have collected rainfall data from several regions of India, including: Mumbai (Maharashtra) Kolhapur (Maharashtra), Chennai (Tamil Nadu), Meghalaya, Arunachal Pradesh, Sikkim, West Bengal, Nagaland. Among these, Meghalaya especially the town Mawsynram records the highest average annual rainfall, making it the wettest place in India and the world, with over 11,000 mm/year. The India Meteorological Department (IMD) is the official national agency for meteorology and weather forecasting in India. It operates under the Ministry of Earth Sciences, Government of India.Official website: https://mausam.imd.gov.in Real-time weather updates (temperature, rainfall, wind, humidity), Rainfall data by region, season, and year, Weather forecasts (daily, weekly, seasonal), Cyclone and storm alerts, Monsoon updates, Climatological data and historical records, Agro-meteorological services for farmers

Maximum rainfall is often analyzed using:

- Extreme Value Theory
- Gumbel Distribution
- Time Series Analysis to estimate return periods (e.g., 100-year rainfall event)

Return periods (T_s) are used to estimate the interval of time between natural hazard occurrences of a certain size and assess the risks associated with hydrological events, climate extremes, structural failures, and seismicity. Despite T_s being widely used, they are characterized by strong misconceptions and ambiguities. Although return periods have been successfully applied to analyze storm surges, high tides, and extreme precipitation, concerns arise when they are used to assess the probabilities of future global mean sea level rise (SLR).Most studies consider SLR return periods in the context of storm surges or high tides, rather than treating SLR as a separate and unique hazard. However, sea level rise due to storm surges or tides is regional and temporary, whereas global SLR is a long-lasting and slow phenomenon.

The return period is defined here as the average time interval (expressed in years) between occurrences of a rainfall event of a given or greater magnitude. The return period denotes a recurrence interval. It is a statistical measure of how often a rainfall event of a certain magnitude is likely to happen. It is important in relating extreme rainfall to normal rainfall. Rainfall with a 10-year return period is expected to happen only every 10 years. A 100-year return period corresponds to such an extreme event that we expect it to occur only every 100 years.

The return period is expressed as:

$$T = \begin{cases} \frac{1}{F} & \text{for } F \leq 0.5 \\ \frac{1}{1-F} & \text{for } F \geq 0.5 \end{cases}$$

where F is the cumulative distribution function.

In the following sections, the study explores different types of extreme climatic events and their defining thresholds, highlighting real-world examples and impacts. It then evaluates the risks these events pose to human life, economic systems, and public health, particularly in the context of a changing climate. The report progresses into statistical modeling techniques used to analyze extreme data, focusing on specialized probability distributions. It further applies threshold-based methods for detailed modeling and compares estimation approaches. Lastly, the study presents a practical application using real-world rainfall data, supported by simulations and return level calculations to better understand and anticipate future extremes.

Chapter 2:
Extreme Climate Events

You can find valuable information about extreme climatic weather events from this source: Extreme Climatic Weather Events - ResearchGate. The term climate describes the weather patterns of a particular region over a longer period, usually 30 years. For the purpose of this study, we referred to various online sources to gather background information on extreme weather events. Notably, we utilized the following ResearchGate publication to understand the classification and characteristics of such events:

2.1 Extreme Weather

Extreme weather events include unexpected, unusual, unpredictable, severe, or unseasonal weather conditions. This means that the weather is at the extremes of the historical distribution. These types of extreme events are based on a location's recorded weather history. These happen with reference to all four seasons. The climatic events include:

- Cold wave, Fog, Snowstorms
- Thunderstorm, and Dust storms
- Heat wave
- Tropical cyclones and Tidal waves
- Floods, Heavy rain, and Landslides
- Droughts

Some extreme weather and climate events have increased in recent decades. The world has warmed, triggering many other changes to the Earth's climate. Over the last 50 years, the world has seen increases in prolonged periods of excessively high temperatures, heavy downpours, and in some regions, severe floods and droughts.

2.2 Major Extreme Weather Events and Their Thresholds

Extreme weather events are becoming increasingly frequent and intense due to climate change. These events include droughts, floods, cyclones, heatwaves, cold waves, wildfires, and coastal flooding—each with scientifically defined thresholds that mark the point at which weather conditions become hazardous and cause significant societal disruption.

- **Droughts**

Drought is a prolonged period of deficient rainfall relative to the statistical multi-year average, leading to water shortages, crop failures, and socioeconomic stress. In India, meteorological drought is defined by the Indian Meteorological Department (IMD) as a situation where the seasonal rainfall is deficient by more than 25% of the long-term average. Agricultural drought focuses on soil moisture deficits, while hydrological drought considers water levels in rivers, lakes, and reservoirs.

One of the most severe droughts in recent years occurred during 2015–2016, affecting over 10 states including Maharashtra, Karnataka, and Andhra Pradesh. These regions experienced rainfall deficits ranging from 25% to 50%, with many districts reporting Standardized Precipitation Index (SPI) values below –1.5, indicating extreme drought conditions. The consequences were dire: widespread crop failures, depletion of groundwater, cattle deaths, and significant distress migration, particularly in rural areas.

Although direct death tolls from drought are not always recorded, the indirect impacts—such as increased rates of farmer suicides, malnutrition, and long-term economic loss—are considerable. The financial impact of the 2015–16 drought is estimated to have exceeded Rs.60,000 crore due to crop loss, employment loss, and relief expenditures. Droughts underscore the urgent need for climate-adaptive agriculture, water conservation measures, and robust drought early-warning systems.

- **Extreme Rainfall Event**

Extreme rainfall refers to unusually high precipitation over a short period, often leading to urban flooding, landslides, and significant damage to infrastructure. In India, extreme rainfall events are defined by daily rainfall exceeding 244.5 mm, as per the Indian Meteorological Department (IMD). Such events are increasingly frequent due to climate change and the intensification of the monsoon cycle.

A landmark event was the Mumbai cloudburst of 2005, where the city received a record-breaking 944 mm of rainfall in just 24 hours, the highest ever recorded in India for a single day. This extreme downpour paralyzed the city, resulting in over 1,000 deaths, massive traffic congestion, power outages, and economic damages exceeding Rs.2,000 crore. Another recent example is the 2018 Kerala floods, where several districts experienced over 300 mm of rainfall in a single day during the peak monsoon season, contributing to one of the worst flood disasters in the state's history.

These events highlight the need for robust urban drainage systems, updated rainfall thresholds, and early warning mechanisms to minimize damage and ensure climate resilience.

- **Floods**

Floods are one of the most frequent and devastating natural disasters in India, often triggered by intense rainfall, cyclones, or rapid snowmelt.They result in the overflow of water onto land that is usually dry, causing widespread disruption to life, property, and the economy. The Indian Meteorological Department (IMD) considers rainfall exceeding 244.5 mm in 24 hours as "extremely heavy", which significantly increases the likelihood of flash floods and riverine flooding.

One of the most catastrophic flood events in recent history was the 2018 Kerala floods. Triggered by unprecedented monsoon rainfall—over 164% of the normal during the month of August—multiple rivers including the Periyar and Bharathapuzha overflowed, submerging vast areas. Over 400 lives were lost, and more than one million people were displaced across the state. The disaster caused infrastructural damage exceeding Rs.30,000 crore, impacting homes, roads, agriculture, and tourism.

Other notable flood events include the Bihar floods (2017) and Assam floods (almost annual), affecting millions. Floods are exacerbated by unplanned urbanization, poor drainage systems, and encroachments on floodplains. The increasing intensity and frequency of flood events underline the importance of resilient infrastructure, early warning systems, river basin management, and sustainable urban planning.

- **Cyclones and Hurricanes**

Cyclones, also known as hurricanes or typhoons in other regions of the world, are intense circular storms that originate over warm tropical oceans and are characterized by strong winds, heavy rain, and low atmospheric pressure. In the Indian context, these storms are classified as tropical cyclones, and the India Meteorological Department (IMD) categorizes them based on wind speed—from "Depression" (31–49 km/h) to "Super Cyclonic Storm" (wind speed exceeding 221 km/h).

India's long eastern coastline, especially the Bay of Bengal region, is highly vulnerable to cyclones.One of the most destructive in recent history was Cyclone Amphan (2020), which reached wind speeds exceeding 180 km/h and struck the states of West Bengal and Odisha. It caused widespread destruction, including infrastructure damage, uprooting of trees, and flooding. Nearly 100 lives were lost, and the estimated financial loss crossed $13 billion (approx. Rs.1 lakh crore), making it one of the costliest cyclones in the Indian subcontinent.

Other significant cyclonic events include Cyclone Fani (2019), Cyclone Hudhud (2014), and the historic 1999 Odisha Super Cyclone, which claimed over 10,000 lives. With rising sea surface temperatures due to climate change, both the frequency and intensity of cyclones are expected to increase. This necessitates improved forecasting, coastal protection infrastructure, and disaster preparedness at the community level.

- **Heatwaves**

Heatwaves are prolonged periods of excessively hot weather, which may be accompanied by high humidity, especially in coastal areas. In India, the India Meteorological Department

(IMD) declares a heatwave when the maximum temperature reaches at least 40°C in plains and 30°C in hilly areas, and is 4.5°C to 6.4°C above the normal temperature. A severe heatwave is declared when the deviation exceeds 6.4°C.

India has experienced an alarming rise in the frequency, duration, and intensity of heatwaves in recent decades, attributed largely to climate change, urban heat island effects, and deforestation. The 2015 heatwave was one of the deadliest in Indian history, claiming over 2,500 lives, particularly in the states of Andhra Pradesh and Telangana. Temperatures soared above 47°C in some regions, overwhelming healthcare systems and causing widespread dehydration, heat strokes, and crop damage.

Heatwaves not only result in significant mortality and health crises but also severely impact agriculture, labor productivity, and energy consumption. The urban poor and outdoor workers are especially vulnerable. With climate models predicting more frequent and intense heatwaves in the coming years, there is a critical need for public awareness campaigns, urban greening strategies, heat action plans, and early warning systems to minimize human and economic losses.

- **Cold Waves**

Cold waves are sudden drops in minimum temperatures over a region, often accompanied by fog and icy winds, leading to extreme cold weather conditions. According to the India Meteorological Department (IMD), a cold wave is declared when the minimum temperature is 4.5°C to 6.4°C below the normal, and a severe cold wave is declared when it is 6.5°C or more below the normal, particularly in northern and central India.

These events typically occur during the winter months, from December to February, and most frequently impact the Indo-Gangetic plains, including states like Uttar Pradesh, Punjab, Haryana, Rajasthan, and Bihar. Vulnerable populations such as the elderly, children, and the homeless are particularly at risk during these events.

A significant cold wave occurred in 2013, resulting in over 100 fatalities across northern India. Prolonged exposure to cold conditions caused hypothermia, respiratory issues, and other health complications. In addition to human health, cold waves can damage rabi crops like wheat and mustard, and lead to livestock mortality, further affecting rural livelihoods.

While often underreported compared to other extreme events, cold waves continue to pose a serious public health concern.The increasing variability of winter temperatures due to climate change highlights the need for improved forecasting, public shelters, and early warning dissemination to reduce fatalities and economic losses.

- **Wildfires**

Wildfires, also known as forest fires, are uncontrolled fires that rapidly spread through vegetation, particularly in dry and drought-prone regions. In India, wildfires are most common during the pre-monsoon summer months (March to May), when high temperatures, low humidity, and dry vegetation create ideal conditions for fire outbreaks. While natural causes like lightning can trigger wildfires, human activities such as shifting cultivation, careless disposal of cigarettes, or deliberate burning are often responsible.

Regions frequently affected by wildfires include Uttarakhand, Himachal Pradesh, Odisha, Madhya Pradesh, and parts of the Western Ghats. A notable recent example is the 2019 Bandipur wildfire in Karnataka, which burned over 10,000 acres of forest. Though human casualties were avoided, the fire caused significant ecological damage, destroyed wildlife habitats, and released large amounts of carbon into the atmosphere.

Unlike floods or cyclones, wildfires often do not result in high human mortality but lead to substantial environmental and economic loss, especially in terms of biodiversity, timber resources, and tourism. In a changing climate, rising temperatures and prolonged dry spells are expected to make Indian forests even more vulnerable to fire. Strengthening fire detection systems, promoting community forest management, and equipping forest personnel with modern firefighting tools are critical steps toward wildfire mitigation.

- **Sea Level Rise and Coastal Flooding**

Sea level rise refers to the gradual increase in the global average sea level, primarily driven by climate change through thermal expansion of seawater and the melting of glaciers and ice sheets. Coastal flooding, on the other hand, is the temporary inundation of low-lying coastal areas due to storm surges, high tides, or heavy rainfall, and its frequency is significantly amplified by rising sea levels.

India, with a coastline of over 7,500 kilometers, is particularly vulnerable to sea level rise. According to studies by the Indian National Centre for Ocean Information Services (INCOIS), sea levels along India's coasts are rising at a rate of about 1.3 mm to 3.3 mm per year. Major cities such as Mumbai, Kolkata, Chennai, and Visakhapatnam face increasing threats from high tide flooding, coastal erosion, and saline water intrusion.

For instance, Chennai experienced severe coastal flooding in 2015 during the northeast monsoon, which was worsened by poor drainage and storm surges. The floods displaced thousands and caused damage exceeding Rs.15,000 crore. Similarly, frequent tidal flooding in the Sundarbans poses a serious threat to agriculture, livelihoods, and the region's unique mangrove ecosystem.

If global warming continues unchecked, projected sea level rise by the end of the century could submerge several low-lying coastal zones in India, displacing millions. Adaptation measures such as mangrove restoration, improved coastal zoning, early warning systems, and climate-resilient infrastructure are essential to safeguard vulnerable coastal communities.

These scientifically defined thresholds are crucial for issuing early warnings, planning emergency responses, and informing long-term policy and infrastructure development. As climate change continues to alter global weather patterns, understanding and monitoring these thresholds becomes essential for minimizing the adverse impacts on both human life and the environment.

In the upcoming sections, the focus shifts to understanding how climate-induced extreme events affect human vulnerability, health, and economic stability, offering insights through statistical indicators like mortality rates. The study then introduces theoretical modeling frameworks using extreme value distributions to characterize rare environmental phenomena. Building on this, threshold-based approaches are implemented to estimate return levels and assess the likelihood of future extremes. The final part of the report showcases a real-world application involving rainfall data, where simulations and statistical models are used to interpret and forecast environmental risk.

Chapter 3:
Risk management of extreme events under climate change

Risk management is an effective way to mitigate the adverse consequences of extreme events, and plays an important role in climate change adaptation. On the basis of the literature, this paper presents a conceptual framework for managing the risk of extreme events under climate change, and accordingly summarizes the recent developments with a focus on several key topics. In terms of risk determinants, the impacts of climate variability on the frequency of extreme events are addressed, and the various meanings and measurements of specific vulnerability are compared. As for the process of risk management, the dynamic assessment approach regarding future climate condition is emphasized. Besides, in view of decision making the available means to enhance the effectiveness of adaptation and mitigation strategies are highlighted. Finally, uncertainty is discussed with respect to its sources and solution.

Climate change may cause serious impacts on human-environmental system, and is an integrated scientific issue which challenges the world Fischedick et al. [2014], It is reported that the changing climate may result in more extreme events worldwide, so that there would be heavier socioeconomic damages Murray and Ebi [2012] ,Rummukainen [2012] ,Yuan et al. [2016]. This is receiving more attention from the public, and especially the governments and research scholars have been devoted to exploring effective measures to mitigate adverse consequences.

Risk management is an available way to timely cope with extreme events Nam et al. [2012]. Different from traditional idea, it aims to emphasize preparedness and provide appropriate strategies according to the extent of damage. In the context of climate change, the occurrence of extreme event and socioeconomic development appear to own high uncertainty with varying time and space. This suggests that risk management is of great significance to help alleviate the impacts of weather-related extremes, and of necessity in adaptation to climate change (IPCC, 2012; Kunreuther et al., 2013).

It is argued that the risk of climate change, which mainly arises from extreme events, reflects the interactions between hazard and vulnerability in a particular condition which integrate natural and social sciences (Blaikie et al., 1994; UN/ISDR, 2004). Thus, risk management of extreme events under climate change is regarded as an interdisciplinary problem, and there have been some discussions in different aspects.

The cause of risk is attributed to hazardous physical event whose variations are expected to influence the components of risk management. With global environmental change, therefore, there are more complicated characteristics of risk management of extreme events, and practically these bring out some bigger challenges. First, it is required to analyze the effects of climate change on extreme events and the associated consequences of human-environmental system. This refers to risk assessment which attempts to describe climate change risk with qualitative and quantitative methods. Second, it needs to detect the ways to set up coping strategies with diverse information and knowledge, and the adoption of adaptive behavior in practice. This relates to damage adaptation and mitigation which intend to reduce and control the risk of extreme events. Finally, the uncertainty should be considered with respect to the possible impacts and solutions because of its essential role in risk management.

This paper aims to highlight the features of climate change risk, and address the advances in risk management. The crucial components in risk management are identified based on a bibliometric analysis. Accordingly, a conceptual framework for risk management of extreme events under climate change is presented to summarize recent developments with a focus on some key topics.

Conceptual Framework

The bibliometric analysis is made with the data collected from Web of Science. On the basis of the literature, a conceptual framework for risk management of extreme events considering climate change effect is given as Figure 1.

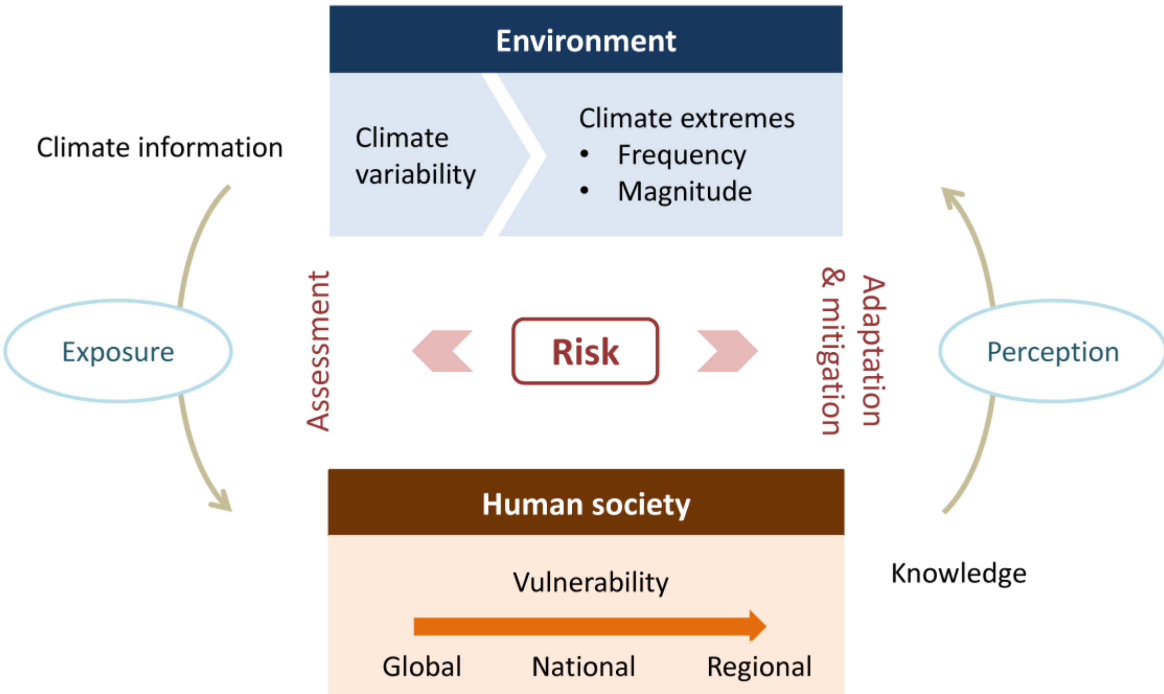

Figure 3.1: Framework for Climate Risk: Interaction Between Environmental Extremes and Human Vulnerability

Conceptual framework for risk management of extreme events under climate change (adapted from Turner et al. (2003); UN/ISDR (2004); IPCC (2012))

For risk management, the basic work is to address how to characterize risk and how to deal with risk. These eventually refer to risk assessment and risk adaptation and mitigation.

The risk of extreme events, which results from the interactions between climate and human society, consists of three primary components including hazard, exposure, and vulnerability (IPCC, 2012). Here hazard refers to various kinds of climate extremes, and its extent is often characterized by frequency and magnitude. Climate variability can directly influence natural environment on both temporal and spatial dimensions, so that there would be changes in the statistical characteristics of extreme events. These affect human society via particular exposure, and thus vulnerability is commonly defined as the degree to which a system is likely to be adversely affected (Adger, 2006). Notably, vulnerability at different scales actually carries diverse information (Fekete et al., 2010). This requires the integration of multi-level information in vulnerability analysis.

Risk assessment synthesizes hazard, exposure, and vulnerability to map risk with qualitative and quantitative methods. Hazard analysis and vulnerability analysis are two basic processes to find the relationship between the extent of extreme event and its probability of occurrence and the relationship between the extent of extreme event and the magnitude of consequence respectively. Therefore, the outcomes of risk assessment have various types. For example, risk classification is a quantitative form that reveals the differences in risk level across areas (Yuan et al., 2015a). This facilitates the exploration of risky nations and regions at the macro-level. Yet, risk curve quantifies the relationship between the probability of occurrence of extreme event and the magnitude of consequence to provide more detailed information for risk description. Due to climate variability and socioeconomic development, the dynamic risk assessment regarding future climate condition is of greater practical significance.

To mitigate and adapt to climate change risk, the structural and non-structural measures are adopted. Structural interventions concern the optimized plan developed by cost-benefit analysis and portfolio according to risk level. As a result, this requires to figure out the acceptable ranges of risk. As for an individual, there are several factors playing key roles in choosing adaptation and mitigation strategies, such as risk preference, risk perception, living experience, and living condition. High risk awareness makes more adaptive behavior such as buying insurance, reducing asset exposure, and preparing emergent facilities, and these could help promote the effectiveness of damage reduction.

Two elements, climate information and knowledge, need to be highlighted in risk management of extreme events under climate change. When making decisions, policy makers, managers, and individuals all rely on climate information which are required to be not only useful but also usable (Lemos et al., 2012). Delivering accurate information could make better strategies and more benefits. In addition, it is argued that human knowledge, which refers to both scientific knowledge and local experience, is necessary during this process. Therefore, it requires more participants with various knowledge and the integration of diverse information to enhance the objectivity of risk assessment and the effectiveness of mitigation and adaptation strategies.

3.1 No.of Deaths and Death Rates Due to Extreme Weather Events

In general, if a phenomenon, such as global warming, could change the frequency, intensity, and/or duration of extreme weather events such as floods, droughts, storms, and extreme temperatures, the result could be an increase in these categories of events in some locations and for some periods and a decrease at other locations and other periods. Some of the effects of these changes will tend to offset each other and/or be redistributed over space and time.

For instance, an increase in deaths due to heat waves at one location might be compensated for by a decline in deaths due to fewer or less intense cold waves at the same or another location. Alternatively, global warming might redistribute the temporal and spatial pattern of rainfall, droughts, and other such events.

Accordingly, to estimate the net impact, if any, of global warming on mortality, it is probably best to examine cumulative deaths at the global level aggregated over all types of extreme weather events. Because of the episodic nature of extreme events, such an examination should ideally be based on several decades, if not centuries, worth of data. Any such examination should, of course, recognize that the quality of the data, data coverage, adaptive capacity, and exposure of human populations to risk also change over time.

In particular, one should examine mortality so as to filter out the effect of population growth on the magnitude of the population at risk. However, use of mortality rates may be insufficient to account for the fact that as the population becomes larger, people will migrate to riskier and more vulnerable locations as the less vulnerable locations are occupied.In addition, inappropriate state policies and the greater availability of insurance place people at moral hazard, encouraging them to bear less than their full burden of financial risk. This may also place even wealthier populations at greater physical risk.

Disasters – from earthquakes and storms to floods and droughts – kill approximately 40,000 to 50,000 people per year. This is the average over the last few decades.While that is a relatively small fraction of all deaths globally, disasters can have much larger impacts on specific populations. Single extreme events can kill tens to hundreds of thousands of people. In the 20^{th} century, more than a million deaths per year were not uncommon. Disasters have other significant impacts, too. Millions of people are displaced – some left homeless – by them each year. The economic costs of extreme events can be severe and difficult to recover from. This is particularly true in lower-income countries, where resources for recovery are limited. We are not defenceless against disasters: deaths from disasters have fallen significantly over the last century as a result of: Early warning systems,Better infrastructure, More productive agriculture,Coordinated emergency responses.These improvements have helped reduce the vulnerability of communities and mitigate the effects of extreme natural events.

Death Rate, also called mortality rate, is a measure of the number of deaths in a population during a specific period.

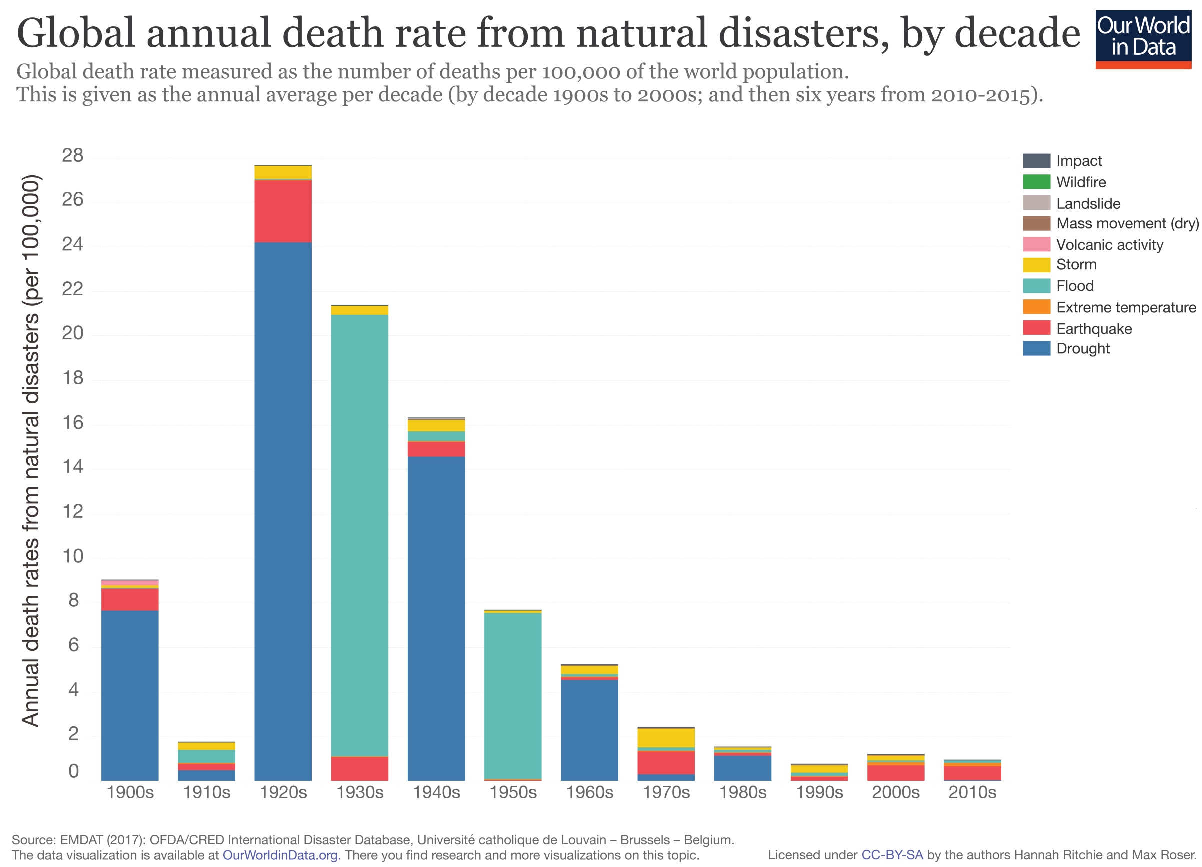

Figure 3.2: Global annual death rate from natural disasters by decade
Source: researchgate.com

$$\text{Death Rate} = \frac{\text{Number of Deaths}}{\text{Total Population}} \times 1,000 \quad (\text{or } 100,000)$$

It is usually expressed as: Deaths per 1,000 people, or Deaths per 100,000 people (for rare events)

Extreme Weather Events: These are unusual, severe, or unseasonal weather conditions, including: Floods,Heatwaves,Cold waves,Cyclones,Droughts,Wildfires,Heavy rainfall or snowfall Causes of Deaths Due to Extreme Weather People may die directly or indirectly because of these events: Direct Causes: Drowning in floods, Heatstroke in heatwaves, Hypothermia in cold waves, Building collapse during storms,Landslides after heavy rain

Indirect Causes: Disease outbreaks after floods (cholera, malaria), Malnutrition due to drought and crop failure, Mental health stress or suicide, Long-term displacement effects

Excess Mortality

$$\text{Excess Deaths} = \text{Observed Deaths} - \text{Expected Deaths}$$

This method estimates how many more people died than usual during an extreme event.

Event-Specific Mortality Rate

$$\text{Event Death Rate} = \frac{\text{Deaths from specific event}}{\text{Exposed Population}} \times 100,000$$

The Heatwave in Europe (2003)

- The summer of 2003 was one of the hottest on record in Europe. The heatwave lasted for several weeks, with temperatures in some areas reaching 40°C (104°F), and in some regions even higher.
- France was the hardest hit, with 14,802 deaths attributed directly to the heat, many of which were elderly individuals who lived alone.
- Hospitals were overwhelmed with patients suffering from heat-related conditions, and public health systems were unable to handle the extreme volume.
- The heatwave led to power outages, droughts, and crop failures, which exacerbated the situation, causing food shortages and increasing the spread of diseases.
- Long-term effects included changes in how countries prepare for future heat events, with improved early warning systems and public health responses.

Country: Multiple European countries, especially France, Spain, Italy, and Germany.
Total Deaths: Estimated at 70,000 deaths.
Cause of Deaths: Heatstroke and dehydration due to high temperatures. Vulnerable populations such as the elderly were the most affected.

The Tsunami in the Indian Ocean (2004)

- On December 26, 2004, a 9.1-9.3 strength earthquake off the coast of Sumatra, Indonesia, triggered a massive tsunami that struck 14 countries around the Indian Ocean, with devastating effects on coastal populations.
- The waves, some of which reached heights of 30 meters (100 feet), swept across parts of Indonesia, Sri Lanka, India, Thailand, and other areas, causing catastrophic loss of life.
- The tsunami hit with little warning and caused widespread devastation. Thousands of people were killed instantly by the waves, while many others perished later from injuries, disease, or lack of access to clean water and food.
- The death rate from the tsunami was staggering, especially in places like Indonesia's Aceh province, where entire villages were destroyed.
- In the aftermath, the international community provided significant aid, and the disaster led to improvements in early warning systems for tsunamis and coastal infrastructure.

Countries Affected: Indonesia, Thailand, Sri Lanka, India, Maldives, and others.
Total Deaths: 230,000-280,000 deaths.
Cause of Deaths: Tsunami triggered by an undersea earthquake, affecting coastal populations. While not a ”weather event,” the tsunami was extreme in its impact, and many deaths occurred due to the extreme force of the waves, coupled with flooding.

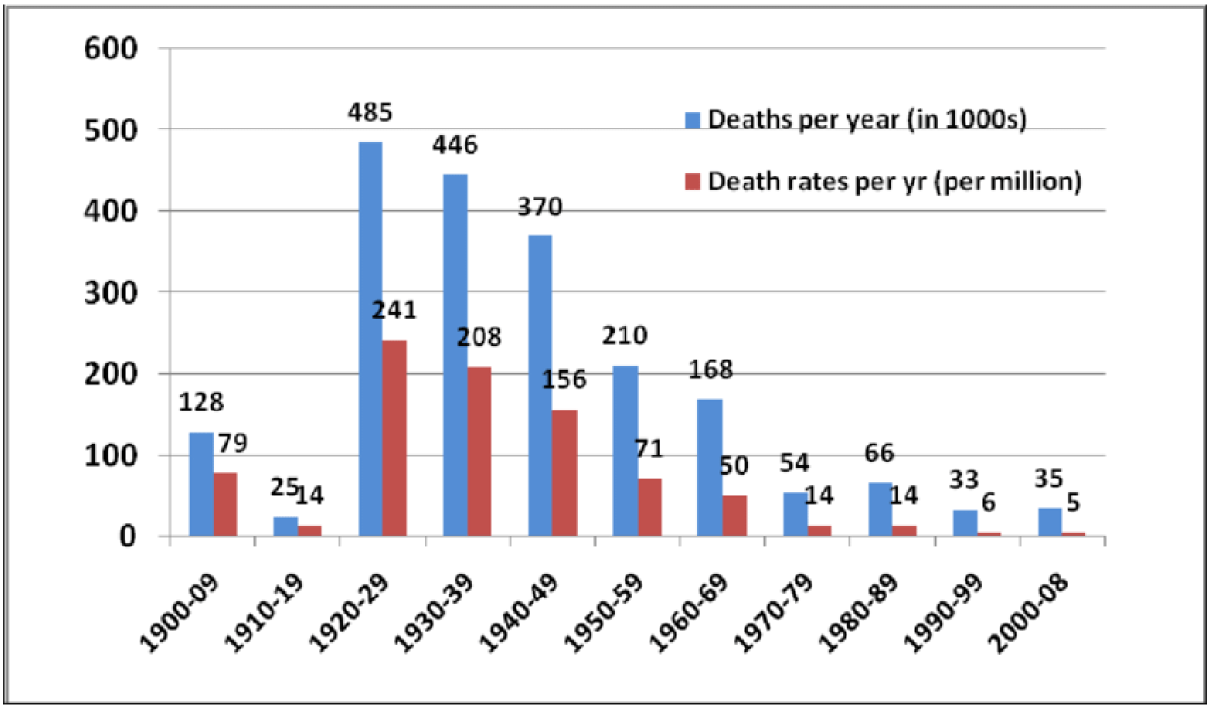

Figure 3.3: Global death and death rates due to extreme events, 1900–2008
Source: researchgate.com

The data reflects an increasing frequency and impact of climate-induced disasters in recent years, emphasizing the urgent need for climate resilience, early warning systems, and disaster preparedness worldwide.

Death Rate Per Year (Annual Death Rate) This measures the number of deaths in a population in a given year, typically expressed per 1,000 people.

$$\text{Death Rate per Year} = \left(\frac{\text{Number of Deaths in a Year}}{\text{Total Population}}\right) \times 1,000$$

Use:General comparison of mortality between countries or over time. Public health planning.

Death Rate Per Million (Per Capita Death Rate) This shows the number of deaths for every one million people in a population, usually used when the death numbers are small or for rare events like natural disasters, pandemics, or extreme weather events.

$$\text{Death Rate per Million} = \left(\frac{\text{Number of Deaths}}{\text{Total Population}}\right) \times 1{,}000{,}000$$

Use: More sensitive measure for low-frequency events.

Table 3.1: Climate-related Disasters and Associated Death Tolls

Year	Event and Location	Deaths
2007	Cyclone Sidr, Bangladesh	4,234
2008	Cyclone Nargis, Southern Myanmar	138,366
2010	Heatwave, Western Russia	55,736
2011	Drought, Somalia	258,000
2013	Floods, Uttarakhand, India	6,054
2013	Typhoon Haiyan, Philippines	7,354
2015	European Heatwave, France	3,275
2022	European Heatwave, Multiple Countries	53,542
2023	European Heatwave, Multiple Countries	37,129
2023	Storm Daniel, Libya	12,352

3.2 The impact of extreme weather events on economy

The economic research on the determinants of budget balances has been going on for several decades. Extreme weather events could affect budget balances both directly and indirectly and could either influence the revenue or the expenditure side.

The direct fiscal impact mainly works via the expenditure side, i.e., via relief payments or the response of the public sector to, for example, human or infrastructure damages. The indirect effects come from the drop in output and the negative wealth effects. Both could lower tax receipts (e.g., for personal and corporate income taxes) and lead to higher expenditure, for example, on unemployment-related payments or other social transfers.

Unfortunately, it is not possible to obtain disaggregated data for the budget balance for such a large country sample in order to test if extreme weather events increase government expenditures or reduce government revenues.

Extreme weather events such as hurricanes, floods, droughts, heatwaves, and wildfires are becoming more frequent and intense due to climate change. These events significantly impact national and local economies through both direct damage and indirect long-term disruptions. Direct Economic Impacts These are the immediate physical damages caused by the weather event: Damage to Infrastructure and Property: Roads, bridges, railways, power lines, schools, hospitals, and homes can be destroyed or damaged. Rebuilding requires huge government spending or insurance payouts.

Example: Hurricane Katrina (2005) caused over $125 billion in damages in the U.S. Destruction of Productive Assets: Factories, warehouses, farms, and businesses often shut down. Machinery and crops may be lost, reducing production capacity. Insurance Costs: Insurers face massive

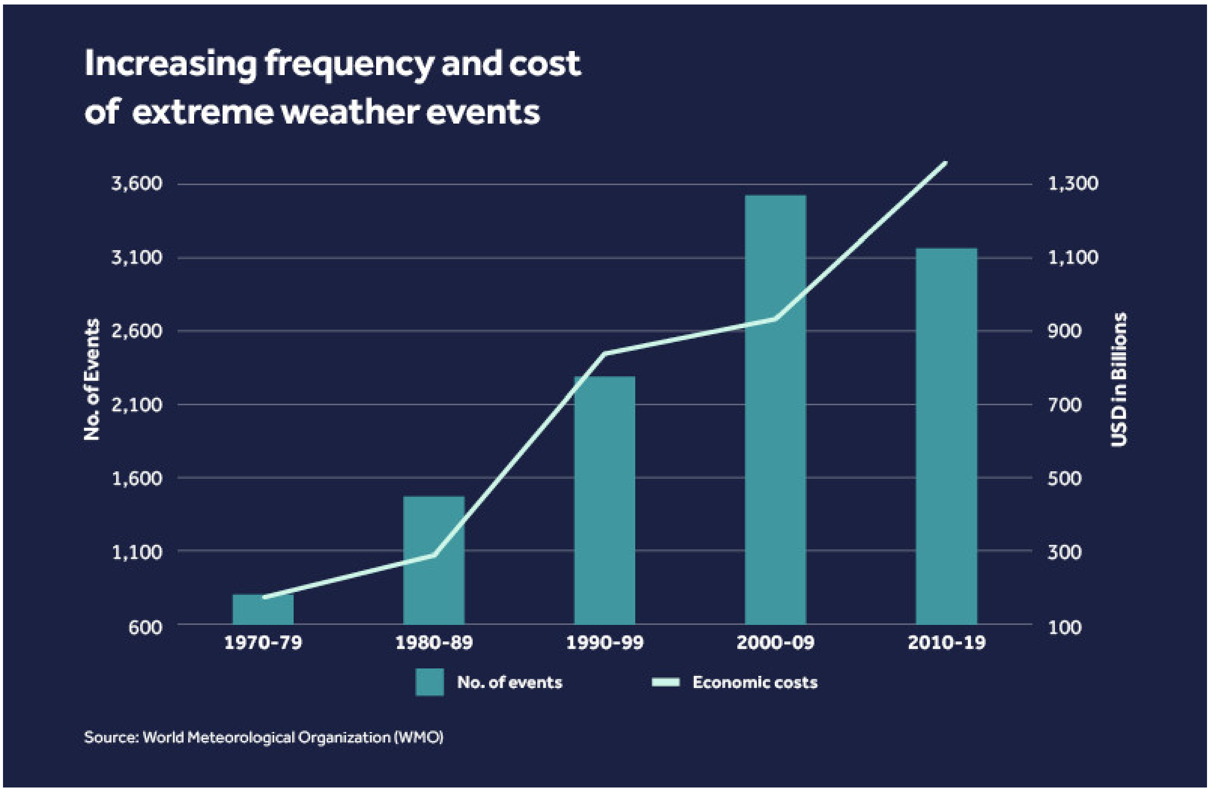

Figure 3.4: The human and economic costs of extreme weather
Source: www.weforum.org

payouts after disasters. This leads to higher insurance premiums for everyone and sometimes the withdrawal of insurance coverage in high-risk areas.

Indirect and Long-Term Economic Impacts These are the longer-term effects on economic systems, often more damaging than the immediate costs: Disruption of Supply Chains: Extreme events delay transportation and cut access to raw materials and products. This affects manufacturing, retail, and exports. Loss of Agricultural Productivity: Droughts, floods, or storms destroy crops and kill livestock. This leads to food shortages, price spikes, and reduced income for farmers. Impact on Labor Productivity: Heatwaves reduce workers' ability to perform manual labor. Health impacts lead to absenteeism and lower work output. Tourism Decline: Areas hit by wildfires, floods, or hurricanes often experience a drop in tourist arrivals. Tourism-dependent economies (like small island states) suffer major losses. Reduced Investment: Investors avoid disaster-prone regions. Damaged investor confidence lowers future economic growth. GDP Loss: The overall Gross Domestic Product (GDP) can decline due to reduced production and consumption. Inflation: Shortages of goods, especially food, lead to rising prices. Debt and Budget Deficits: Governments spend more on disaster response, recovery, and rebuilding, increasing debt. Widening Inequality: Poor communities are more exposed and have fewer resources to recover. This can worsen income inequality and deepen poverty.

3.3 Climate change impacts on health

Climate change presents a fundamental threat to human health. It affects the physical environment as well as all aspects of both natural and human systems – including social and economic conditions and the functioning of health systems. It is therefore a threat multiplier, undermining and potentially reversing decades of health progress. As climatic conditions change, more frequent and intensifying weather and climate events are observed, including storms, extreme heat, floods, droughts and wildfires. These weather and climate hazards affect health both directly and indirectly, increasing the risk of deaths, noncommunicable diseases, the emergence and spread of infectious diseases, and health emergencies.

Climate change is also having an impact on our health workforce and infrastructure, reducing capacity to provide universal health coverage (UHC). More fundamentally, climate shocks and growing stresses such as changing temperature and precipitation patterns, drought, floods and rising sea levels degrade the environmental and social determinants of physical and mental health. All aspects of health are affected by climate change, from clean air, water and soil to food systems and livelihoods. Further delay in tackling climate change will increase health risks, undermine decades of improvements in global health, and contravene our collective commitments to ensure the human right to health for all.

The Intergovernmental Panel on Climate Change's (IPCC) Sixth Assessment Report (AR6) concluded that climate risks are appearing faster and will become more severe sooner than previously expected, and it will be harder to adapt with increased global heating.

It further reveals that 3.6 billion people already live in areas highly susceptible to climate change. Despite contributing minimally to global emissions, low-income countries and small island developing states (SIDS) endure the harshest health impacts. In vulnerable regions, the death rate from extreme weather events in the last decade was 15 times higher than in less vulnerable ones.

Climate change is impacting health in a myriad of ways, including by leading to death and illness from increasingly frequent extreme weather events, such as heatwaves, storms and floods, the disruption of food systems, increases in zoonoses and food-, water- and vector-borne diseases, and mental health issues. Furthermore, climate change is undermining many of the social determinants for good health, such as livelihoods, equality and access to health care and social support structures. These climate-sensitive health risks are disproportionately felt by the most vulnerable and disadvantaged, including women, children, ethnic minorities, poor communities, migrants or displaced persons, older populations, and those with underlying health conditions.

In the following sections, the study moves toward the statistical modeling of extreme environmental data. It begins with an exploration of key extreme value distributions and their theoretical foundations, which are used to model rare and impactful events. Building on this, the report applies threshold-based methods such as the Peaks Over Threshold approach to estimate parameters and return levels using real-world data. These methods are then brought together in a practical application, where actual rainfall data is analyzed and simulations are used to assess the frequency and severity of extreme occurrences, offering valuable insights for future risk prediction and mitigation.

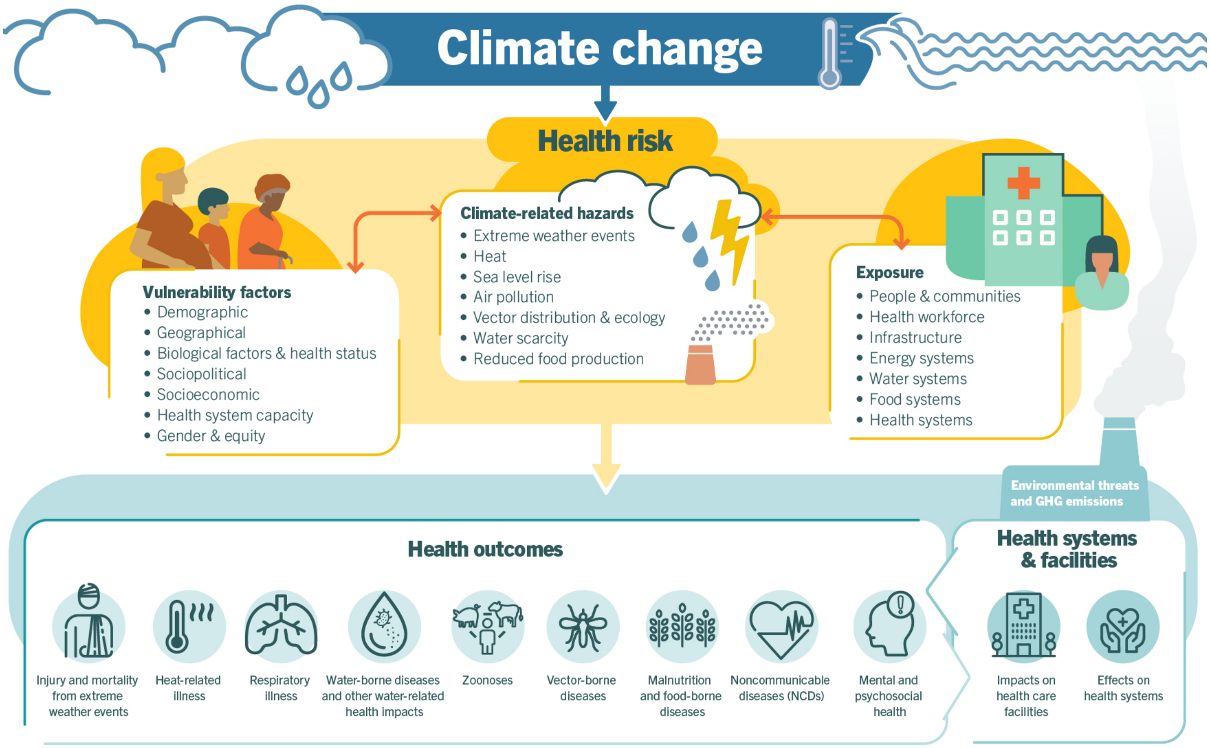

Figure 3.5: An overview of climate-sensitive health risks, their exposure pathways and vulnerability factors. Climate change impacts health both directly and indirectly, and is strongly mediated by environmental, social and public health determinants.

source:www.who.int

Chapter 4:
Parametric Families in Extreme Value Distributions

Probabilistic Extreme Value Theory examines the behavior of extreme events such as maximum and minimum values across various domains, including meteorology, engineering, and finance. The theory traces its roots to early 18th-century probability studies, with significant contributions from Gumbel, Fisher, Tippett, von Mises, and Gnedenko.

EVT focuses on the tails of distributions, enabling specialized statistical models for rare but impactful occurrences like floods, material failure, and financial crashes. Initially applied in astronomy and engineering, it later gained attention in hydrology, climatology, and risk analysis. The theoretical foundations were formalized through asymptotic limit laws, leading to practical applications in diverse fields.

Key works by Gumbel [1958], Galambos [1995] and Davis and Resnick [1988] have advanced both the theoretical and applied aspects of EVT, making it a crucial tool in statistical modeling of extreme events.

4.1 EVD: Theoretical Foundation

The theory of EVD emerged from early 20th-century work by Fisher and Tippett, who showed that the distribution of block maxima from i.i.d. samples converges to one of three types. These were later popularized by Emil Gumbel for practical applications in fields like hydrology and meteorology. EVDs are classified into three primary types based on the tail behavior of the data: Type I (Gumbel) for light-tailed distributions, Type II (Fréchet) for heavy-tailed distributions, and Type III (Weibull) for bounded distributions. The following sections provide a detailed explanation of each type and its applications.

4.1.1 Gumbel Distribution (EVD Type I)

The Gumbel distribution was introduced by Emil Julius Gumbel in the 1940s as part of his pioneering work in the field of extreme value theory. His 1958 book, *Statistics of Extremes*Gumbel [1958], helped establish the Gumbel distribution as a key model for analyzing extreme events in natural and applied sciences.

The Gumbel distribution, also known as the EVD Type I, is used to model the distribution of the maximum (or minimum) values from datasets with light, exponentially decaying tails. It is a special case of the Generalized Extreme Value distribution where the shape parameter is equal to zero. The distribution is especially suitable for modeling extremes in datasets where very large or very small values are possible but not overly frequent.

Common applications of the Gumbel distribution include the modeling of extreme weather events such as maximum daily temperatures or flood levels, as well as stress testing in finance and reliability analysis in engineering. It is particularly useful for studying maximum values drawn from large samples, where the tails of the distribution decay at an exponential rate.

The PDF of the Gumbel distribution is given by:

$$f(x; \mu, \sigma) = \frac{1}{\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)} e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}, \quad x \in \mathbb{R}.$$

The CDF is expressed as:

$$F(x; \mu, \sigma) = e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}, \quad x \in \mathbb{R}.$$

The Gumbel distribution is defined by two key parameters:

- Location parameter (μ): This determines the center of the distribution and shifts it along the x-axis.
- Scale parameter ($\sigma > 0$): This controls the spread of the distribution, affecting the variability of extreme values. A larger σ results in a wider distribution.

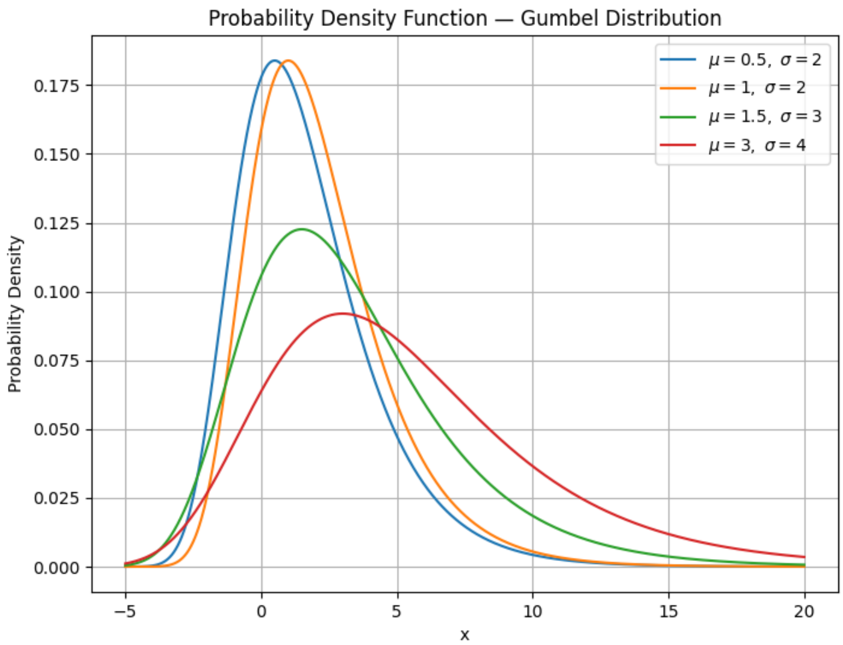

Figure 4.1: Gumbel Type I: PDF with Varying Parameters

The Gumbel distribution is widely used in extreme value theory and risk assessment, providing valuable insights into rare and extreme occurrences in various domains. Its ability to model extreme deviations makes it an essential tool in statistics for analyzing and predicting extreme events.

4.1.2 Fréchet Distribution (EVD Type II)

The Fréchet distribution, named after French mathematician Maurice Fréchet, is one of the three classical types of Extreme Value Distributions. It was introduced in the early 20th century and is particularly useful for modeling extreme values in datasets where large outliers are likely. The Fréchet distribution belongs to the family of extreme value distributions, alongside the Gumbel, Weibull, and GEV distributions.

It is used to model the maximum values in datasets with heavy-tailed characteristics, where extremely high observations occur with non-negligible probability. This makes it especially valuable in fields such as hydrology, finance, environmental science, survival analysis, and engineering. Applications include modeling flood peaks, extreme market losses, and the longest lifespans in survival studies.

The key feature of the Fréchet distribution is its heavy right tail, which enables it to account for rare but significant extreme events that lighter-tailed distributions might underestimate.

The PDF of the Fréchet distribution is given by:

$$f(x;\mu,\sigma,\xi) = \frac{\xi}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{-1-\xi} e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}}, \quad x > \mu.$$

The CDF is expressed as:

$$F(x;\mu,\sigma,\xi) = \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}\right], \quad x > \mu.$$

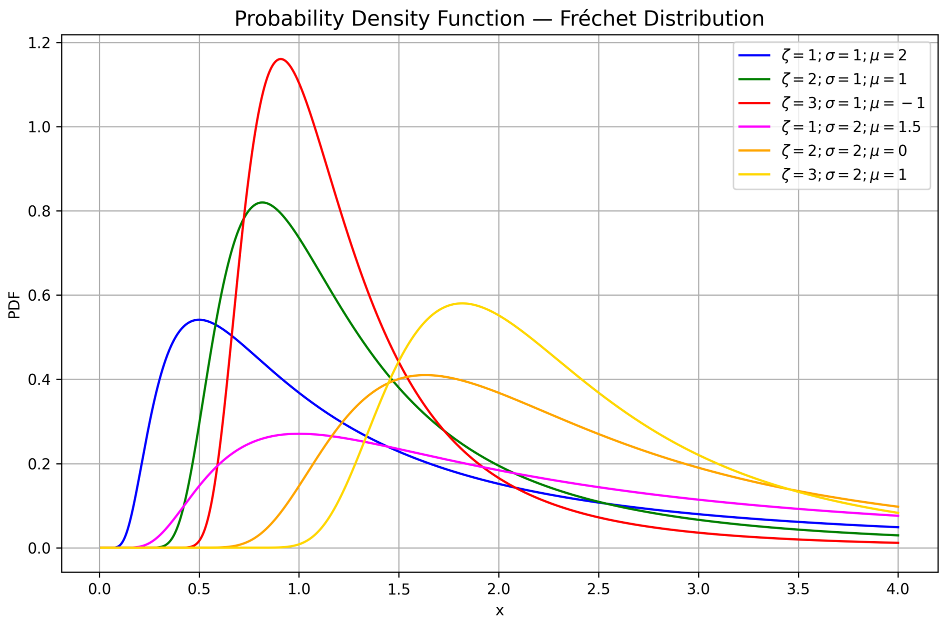

Figure 4.2: Fréchet Type II: PDF with Varying Parameters

The Fréchet distribution is defined by three key parameters:

- Location parameter (μ): This determines the minimum threshold beyond which the distribution applies. It shifts the distribution along the x-axis.

- Scale parameter ($\sigma > 0$): This controls the spread of the distribution, affecting the range of extreme values. A larger σ results in a wider distribution.

- Shape parameter ($\xi > 0$): This governs the heaviness of the right tail. A larger ξ indicates a heavier tail, meaning that extreme values occur more frequently.

Due to its flexibility, the Fréchet distribution is highly valuable in extreme value analysis. By adjusting its parameters, it can effectively model different datasets, helping in risk assessment, planning for rare events, and making predictions in various domains.-

4.1.3 Weibull Distribution (EVD Type III)

The Weibull distribution, named after Swedish engineer Waloddi Weibull, is a classical model introduced in the mid-20th century and is now widely used in reliability engineering and survival analysis. As the Type III Extreme Value Distribution (EVD-III), it is primarily employed to model the minimum values in datasets, particularly those involving time-to-failure or life durations.

In contrast to the Gumbel and Fréchet distributions, which focus on extreme maxima, the Weibull distribution is suited for analyzing the minimum or weakest values in a sample. This makes it particularly useful in applications such as modeling the life expectancy of components, identifying the earliest failure points in systems, or studying the shortest survival times in medical research. Its flexibility and relevance to real-world failure mechanisms have made it a standard tool in risk analysis and reliability modeling.

The PDF of the Weibull distribution is given by:

$$f(x; \mu, \sigma, \xi) = \frac{\xi}{\sigma} \left(\frac{x-\mu}{\sigma} \right)^{\xi - 1} e^{-\left(\frac{x-\mu}{\sigma}\right)^{\xi}}, \quad x > \mu.$$

The CDF is expressed as:

$$F(x; \mu, \sigma, \xi) = 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)^{\xi}}, \quad x > \mu.$$

The Weibull distribution is characterized by three key parameters:

- Location parameter (μ): Determines the minimum threshold value of the distribution. It shifts the distribution along the x-axis.

- Scale parameter ($\sigma > 0$): Controls the spread of the distribution. A larger σ results in a wider range of possible values.

- Shape parameter ($\xi > 0$): Governs the distribution's shape, affecting the probability of extreme minimum values. Different values of ξ can result in different tail behaviors, making the Weibull distribution highly flexible.

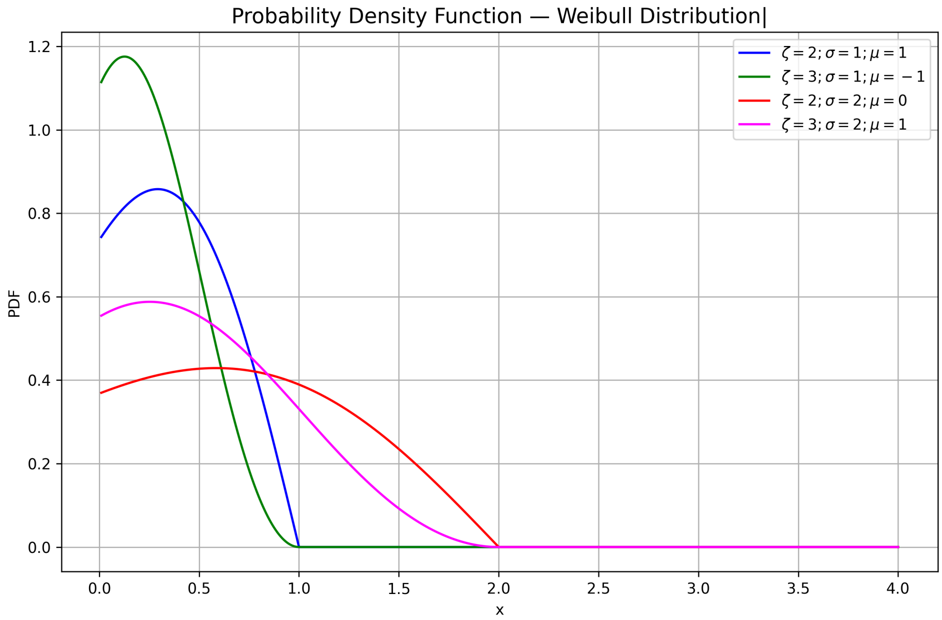

Figure 4.3: Weibull Type III: PDF with Varying Parameters

The Weibull distribution is a powerful tool in statistics for analyzing extreme minimum values across different disciplines. It is widely used in predicting failure rates of materials, estimating product life cycles, and assessing risks in environmental and financial applications. Due to its adaptability, it plays a crucial role in decision-making and risk assessment.

4.2 Generalized Extreme Value Distribution

The GEVD unifies the three types of Extreme Value Distributions: Gumbel (Type I), Fréchet (Type II), and Weibull (Type III). It models the distribution of block maxima and is widely used in extreme value analysis across various fields.

The PDF of GEVD is given by:

$$f(x; \mu, \sigma, \xi) = \frac{1}{\sigma}\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}-1} \exp\left[-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right],$$

where the condition $1 + \xi\frac{x-\mu}{\sigma} > 0$ must hold.

The cumulative distribution function (CDF) of GEVD is given by:

$$F(x; \mu, \sigma, \xi) = \exp\left(-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right)$$

for $1 + \xi\frac{x-\mu}{\sigma} > 0$.

- μ (location parameter): Determines the center of the distribution.
- $\sigma > 0$ (scale parameter): Controls the spread of the distribution.
- ξ (shape parameter): Governs the tail behavior.
  If $\xi > 0$, the distribution follows a Fréchet-type heavy-tailed form.
  If $\xi = 0$, it reduces to the Gumbel distribution.
  If $\xi < 0$, it follows a Weibull-type bounded distribution.

The GEVD is a powerful tool for modeling extreme environmental events. In meteorology and climatology, it is used to analyze extreme weather events, such as maximum rainfall, record high wind speeds, and temperature extremes, all of which are crucial for understanding climate change impacts and preparing for natural disasters. In hydrology, GEVD is applied to predict extreme flood levels and river discharge events, which are critical for designing infrastructure like dams and levees to mitigate flood risks. GEVD also plays a vital role in assessing environmental hazards such as air pollution peaks, extreme storm events, and other climate-induced threats. By modeling these extreme events, GEVD helps improve environmental planning, risk mitigation strategies, and the overall understanding of how environmental conditions may evolve under changing climatic patterns.

The importance of GEVD lies in its ability to unify the three extreme value distributions (Gumbel, Fréchet, and Weibull) under a single framework. By adjusting its shape parameter, it can flexibly model different types of extreme events, making it a powerful tool in extreme value theory. Its ability to capture rare and catastrophic occurrences allows researchers and practitioners to make informed decisions, develop robust safety measures, and optimize resource allocation in various industries. Understanding and implementing GEVD effectively can significantly improve risk assessment strategies and resilience planning against extreme and unpredictable phenomena.

4.3 Other Distributions for Extreme Events

Beyond the GEVD, there are several other parametric families that help in modeling extreme events with more complex dependencies and characteristics. These include bivariate and multivariate distributions, circular models for directional data, and time series extensions that account for temporal dependencies. The following sections highlight some of these distributions and their applications in various fields.

Generalized Pareto Distribution (GPD)

The GPD plays a central role in modeling the tail behavior of a distribution, especially when analyzing the excesses over a threshold. It is a key component of the Peak Over Threshold (POT) method, widely applied in fields like hydrology, finance, and environmental sciences.

The GPD offers a flexible framework to handle both heavy-tailed and light-tailed data, making it invaluable for predicting rare but impactful events such as extreme floods, market crashes, or high pollution levels.

Bivariate Logistic Distribution

This distribution is used to model the dependence between two extreme variables that share logistic marginal distributions. It is particularly effective when the variables tend to behave similarly under extreme conditions. The bivariate logistic is suitable for situations where joint extremes are of interest, such as simultaneous extreme rainfall in two regions or joint financial downturns in connected markets.

Bi-logistic Distribution

A special symmetric case of the bivariate logistic, the bi-logistic distribution assumes that both extreme variables exhibit similar behaviors and are likely to reach high values together. It is particularly useful when modeling symmetric joint extremes, where the extremes occur in a balanced and uniform manner between the variables.

Bivariate Negative Logistic Distribution

This distribution is designed to capture asymmetric dependence between two extreme values, where one variable is more likely to experience an extreme event than the other. It is particularly relevant in finance and environmental sciences, where such asymmetry in extreme behavior is often observed, for example, when losses in one asset class do not necessarily imply losses in another.

Circular Distributions

Circular distributions are used for data that naturally wrap around a circle, such as angles, directions, or time-of-day measurements. They are especially relevant in environmental and atmospheric studies for modeling phenomena like wind direction, wave orientation, or the timing of seasonal weather events. These distributions are effective in capturing patterns and extremes associated with directional data.

Time Series Logistic Distribution

This distribution extends the logistic distribution to handle time-dependent data. It is valuable for modeling extremes in sequential datasets, such as peak temperatures, high wind speeds, or extreme financial returns observed over time. It accounts for temporal dependence, making it suitable for studying how extreme events evolve and cluster over time.

Dirichlet Distribution

The Dirichlet distribution is commonly used to model proportions or categorical outcomes, making it useful in multivariate extreme value analysis where observations are distributed across multiple categories or regions. It finds applications in environmental monitoring, insurance claim distributions, and risk assessments involving multiple components or regions.

Each of these distributions plays a unique role in understanding and modeling joint and multivariate extremes, providing critical insights into the behavior of rare and impactful events across various domains such as finance, hydrology, engineering, and environmental sciences.

The upcoming sections extend the theoretical groundwork into practical application by introducing the Generalized Pareto Distribution (GPD) and the Peaks Over Threshold (POT) approach, which are powerful tools for modeling extreme events beyond a selected threshold. These methods allow for more precise estimation of risk measures and return levels. The final part of the study demonstrates the implementation of these techniques using real rainfall data, offering simulation-based analysis and case studies to showcase how statistical modeling can effectively support environmental risk management and decision-making.

Chapter 5:
Generalized Pareto Distribution & POT Approach

The Peaks Over Threshold method emerged as an important development in EVT to model exceedances over a high threshold, rather than block maxima. Its foundation lies in the Pickands–Balkema–de Haan theorem (1974–1975), which showed that for a large class of distributions, the distribution of exceedances converges to a GPD as the threshold increases. The GPD was introduced by John Pickands III in 1975 PICKANDs III [1975], providing a flexible tool to capture the tail behavior of distributions. This approach is now widely used in climate studies, finance, hydrology, and engineering for modeling rare and extreme events beyond a fixed threshold.

In the analysis of extreme events like floods, heatwaves, or market crashes, traditional models often fail to capture the behavior of rare observations. The POT approach in EVT addresses this by focusing on exceedances over a high threshold. The GPD is commonly used to model these exceedances, offering flexibility in capturing various tail behaviors. It helps estimate the probability and magnitude of rare events, supporting applications in hydrology, finance, meteorology, and environmental science.

5.1 Statistical Properties of the GPD

The GPD is widely used in modeling exceedances over a high threshold, particularly in the Peaks Over Threshold (POT) approach in Extreme Value Theory. It is defined by three parameters: the shape parameter ξ, the scale parameter $\sigma > 0$, and a location parameter μ (typically set to the threshold value u).

The PDF of the GPD is given by:

$$f(x) = \begin{cases} \frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{\xi}}, & \text{if } \xi \neq 0,\ x \geq \mu,\ \text{and } \left(1+\xi\frac{x-\mu}{\sigma}\right) > 0, \\ \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right), & \text{if } \xi = 0,\ x \geq \mu. \end{cases}$$

The CDF is expressed as:

$$F(x) = \begin{cases} 1-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1-\exp\left(-\frac{x-\mu}{\sigma}\right), & \text{if } \xi = 0. \end{cases}$$

- Shape parameter (ξ): Controls the tail behavior.
  $\xi > 0$: Heavy tail (e.g., Pareto-type behavior).
  $\xi = 0$: Exponential distribution (light tail).
  $\xi < 0$: Bounded upper tail.

- Scale parameter ($\sigma > 0$): Determines the spread of the distribution.
- Location parameter (μ): Represents the minimum value or threshold; typically chosen based on the problem context.

The domain (support) of the GPD depends on the value of the shape parameter ξ:

$$x \in \begin{cases} [\mu, \infty), & \text{if } \xi \geq 0, \\ \left[\mu, \mu - \frac{\sigma}{\xi}\right), & \text{if } \xi < 0. \end{cases}$$

- **Mean & Variance:**

$$\mathbb{E}[X] = \mu + \frac{\sigma}{1-\xi}, \xi < 1$$

$$\mathrm{Var}(X) = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}, \xi < \frac{1}{2}$$

It is important to note that the existence of moments depends on the value of ξ. As ξ increases, fewer moments exist, which reflects the increasing heaviness of the tail — a critical aspect in modeling extreme events.

5.2 Parameter Estimation Using MLE

Let $x_1, x_2, \ldots, x_n$ be a random sample of independent observations representing exceedances over a high threshold μ. The goal is to estimate the shape parameter ξ and the scale parameter σ of the Generalized Pareto Distribution using the method of MLE.

Likelihood Function

Assuming μ is fixed (as the threshold), the likelihood function for the GPD parameters based on the observed exceedances is:

$$L(\xi, \sigma \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sigma} \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{-1-\frac{1}{\xi}}, \quad \text{for } \xi \neq 0.$$

The corresponding log-likelihood function is:

$$\ell(\xi, \sigma) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log \left(1 + \xi \frac{x_i - \mu}{\sigma}\right).$$

To ensure the density is well-defined, we impose the constraint:

$$1 + \xi \frac{x_i - \mu}{\sigma} > 0 \quad \text{for all } i = 1, 2, \ldots, n.$$

In the case where $\xi = 0$, the distribution reduces to the exponential distribution with parameter σ, and the log-likelihood becomes:

$$\ell(\sigma) = -n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^{n} (x_i - \mu).$$

The MLE of σ in this case is:

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu).$$

In general, there is no closed-form solution for the MLEs of ξ and σ when $\xi \neq 0$. Therefore, numerical optimization methods such as the Newton-Raphson method or built-in functions in statistical software (e.g., `fitdistrplus` in R or `scipy.stats.genpareto.fit` in Python) are used to obtain the estimates.

Under regularity conditions, the MLEs of ξ and σ are consistent and asymptotically normal. That is,

$$\begin{pmatrix} \hat{\xi} \\ \hat{\sigma} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} \xi \\ \sigma \end{pmatrix}, \; I^{-1}(\xi, \sigma) \right),$$

where $I^{-1}(\xi, \sigma)$ is the inverse of the Fisher information matrix.

Accurate parameter estimation is crucial in extreme value analysis, as the tail behavior determined by ξ directly influences risk measures such as return levels and exceedance probabilities.

5.3 Return Levels and Risk Measures

In extreme value analysis, especially using the POT approach with the GPD, return levels are used to quantify the magnitude of rare events expected to occur once every m time units (e.g., years, days). These are crucial in risk assessment and decision-making in fields such as hydrology, finance, and environmental science.

Definition of Return Level

Let z_m denote the return level associated with a return period of m observations. It is defined as the level exceeded on average once every m observations (or time units). Mathematically,

$$\mathbb{P}(X > z_m) = \frac{1}{m}.$$

Return Level Formula under GPD

Assuming that exceedances over a threshold μ follow a Generalized Pareto Distribution, the return level z_m is given by:

$$z_m = \mu + \frac{\sigma}{\xi}\left((m\lambda)^{\xi} - 1\right), \quad \text{for } \xi \neq 0,$$

$$z_m = \mu + \sigma \log(m\lambda), \quad \text{for } \xi = 0,$$

where:

- μ: Threshold value used in the POT method,
- σ, ξ: Scale and shape parameters of the GPD,
- λ: Empirical exceedance rate, defined as the proportion of observations exceeding the threshold:
  $$\lambda = \frac{k}{n},$$
  where k is the number of exceedances and n is the total number of observations.

The return level z_m gives the magnitude of an event that is expected to be exceeded once every m time units. For example, in environmental studies, a 100-year return level refers to a flood level that is expected to be exceeded once in 100 years on average.

Risk Measures Based on GPD

Two common risk measures derived from the GPD model include:

- **Exceedance Probability:**
  $$\mathbb{P}(X > x \mid X > \mu) = \left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}, \quad \text{for } \xi \neq 0.$$
- **Value-at-Risk (VaR):** In finance and insurance, VaR at confidence level $1 - \alpha$ can be interpreted as the return level for $m = 1/\alpha$, i.e.,
  $$\text{VaR}_{1-\alpha} = z_{1/\alpha}.$$
- **Expected Shortfall (ES):** Also known as Conditional Tail Expectation, this measures the expected exceedance above the VaR and for GPD is given (for $\xi < 1$) as:
  $$\text{ES}_{1-\alpha} = \frac{\text{VaR}_{1-\alpha}}{1-\xi} + \frac{\sigma - \xi\mu}{1-\xi}.$$

These return levels and risk measures provide actionable insights in quantifying and preparing for extreme events, allowing for informed decision-making and policy formulation in risk-sensitive domains.

5.4 GEVD versus GPD

The GEVD is used for analyzing block maxima, focusing on the maximum values within specific intervals or blocks, such as annual or monthly extremes. In contrast, the GPD is applied in the POT method, modeling values that exceed a certain threshold, such as extreme weather events or rare phenomena. The key distinction lies in that GEVD is used for block-based maxima analysis, while GPD is suited for modeling exceedances above a threshold.

Table 5.1: Comparison of GEVD and GPD

Feature	$X \sim$ **GEVD**(μ, σ, ξ)	$X \sim$ **GPD**(μ, σ, ξ)
Description	Models block maxima (or minima) over fixed periods. It describes the limiting distribution of sample maxima, often used in block maxima approach.	Models threshold exceedances. It describes the distribution of values exceeding a high threshold, used in Peaks Over Threshold (POT) analysis.
Parameters	μ (location): Center of distribution of block maxima. $\sigma > 0$ (scale): Dispersion of maxima. ξ (shape): Determines tail behavior.	μ (threshold): Threshold level for exceedance. $\sigma > 0$ (scale): Spread of excesses. ξ (shape): Tail heaviness.
CDF	$G(x) = \exp\left\{-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}$	$H(x) = 1 - \left[1 + \xi\frac{x-\mu}{\sigma}\right]^{-1/\xi}$
Limit as $\xi \to 0$	**Gumbel**: $G(x) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}$	**Exponential**: $H(x) = 1 - \exp\left(-\frac{x-\mu}{\sigma}\right)$
If $\xi > 0$	**Fréchet** distribution (heavy-tailed)	**Pareto** distribution (heavy-tailed)
If $\xi < 0$	**Weibull** distribution (finite upper bound)	**Beta** type distribution (finite upper bound)
Return Level	Return level z_p is the value exceeded on average once every $1/p$ periods. Solve for z_p such that: $$G(z_p) = 1 - p$$	Return level z_m is the value exceeded on average once every m observations. Let $\lambda = \mathbb{P}(X > \mu)$, then: $$z_m = \begin{cases} \mu + \frac{\sigma}{\xi}\left((m\lambda)^{\xi} - 1\right), & \xi \neq 0 \\ \mu + \sigma\log(m\lambda), & \xi = 0 \end{cases}$$

5.5 R Packages for Modeling GPD

This section outlines widely used `R` packages developed for modeling and analyzing extreme events, with a particular focus on the GPD and the GEVD. These distributions form the core of Extreme Value Theory, commonly applied through the POT approach for GPD and the block maxima method for GEVD. The selected packages offer tools for parameter estimation, threshold selection, diagnostic checking, and visualization, using techniques such as Maximum Likelihood Estimation and Bayesian inference. Each package varies in terms of scope and user flexibility, allowing adaptation to different data types and analysis requirements. The discussion includes widely used packages such as `evd`, `ismev`, `evir`, `extRemes`, and `POT`, with a summary of their key features, core functions for extreme value modeling, and relevant authorship information.

Table 5.2: Summary of R Packages for Extreme Value Analysis

Package	Author(s), Year	Key Functions	Purpose
`evir`	McNeil & Pfaff (2000/2001)	`gpd()`, `pot()`, `gpd.plot()`	Toolkit for initial GPD fitting and visualization using POT method. Educational and exploratory use.
`evd`	Alec G. Stephenson (2002)	`fpot()`, `gpd()`	Fit GPD and GEVD models. Useful for threshold exceedances and block maxima modeling.
`ismev`	Heffernan, Tawn & Coles (2001)	`gpd.fit()`, `gpd.diag()`	Classical approach for univariate EVA, provides diagnostic plots and goodness-of-fit tests for GPD models.
`POT`	Stéphane Ribatet (2011)	`fitgpd()`, `fit()`	Full POT modeling framework including GPD fitting, threshold selection, and declustering methods. Advanced EVA functionalities.
`extRemes`	Gilleland & Katz (2016)	`fevd()`	Comprehensive modeling of GPD and GEVD using MLE or Bayesian inference. Provides tools for return levels, diagnostics, and uncertainty estimation.

- **fpot()**: Fits the POT model for threshold exceedances and estimates GPD parameters; from `extRemes` package.
  Syntax: `result <- fpot(data, threshold)`
  Output: GPD parameter estimates (scale and shape), standard errors, and diagnostics.

- **gpd.fit()**: Fits the GPD to univariate data using MLE; from `ismev` package.
  Syntax: `fit <- gpd.fit(data, threshold)`
  Output: Parameter estimates, standard errors, and log-likelihood.

- **gpd.diag()**: Produces diagnostic plots for GPD model fit; from `ismev` package.
  Syntax: `gpd.diag(fit)`
  Output: Return level plot, QQ plot, and other diagnostic visuals.

- **gpd()**: Fits the GPD to threshold exceedances using POT; from `evir` package.
  Syntax: `result <- gpd(data, threshold)`
  Output: GPD parameter estimates with basic diagnostics.

- **pot()**: Identifies and summarizes threshold exceedances; from `evir` package.

Syntax: `result <- pot(data, threshold)`
Output: Exceedance values and summary statistics.

- **gpd.plot()**: Generates visual diagnostics for fitted GPD model; from `evir` package.
  Syntax: `gpd.plot(gpd_object)`
  Output: QQ plot, return level plot, and other checks.

- **fevd()**: Fits GP or GEV distributions using MLE or Bayesian methods; from `extRemes` package.
  Syntax: `model <- fevd(data, type = "GP", threshold = threshold)`
  Output: Parameter estimates, confidence intervals, and return level predictions.

- **fitgpd()**: Fits the GPD using MLE or PWM; from `POT` package.
  Syntax: `fit <- fitgpd(data, threshold, est = "mle")`
  Output: Parameter estimates, confidence intervals, and diagnostic plots.

- **fit()**: General-purpose function for fitting GPD and other extreme value models; from `POT` package.
  Syntax: `fit(data, threshold, model = "gpd")`
  Output: Parameter estimates with model diagnostics and plots.

- **gpd()**: Directly fits GPD model with more control than `fpot()`; from `POT` package.
  Syntax: `result <- gpd(data, threshold)`
  Output: GPD parameter estimates with diagnostic summaries.

The next section focuses on the practical implementation of the Generalized Pareto Distribution (GPD) through simulation-based analysis and real-world applications. Using rainfall data from high-precipitation regions, it demonstrates how GPD models are fitted, validated, and used to estimate return levels. This hands-on approach provides clear insight into the usefulness of extreme value modeling for environmental planning and supports data-driven decision-making in the context of climate risk and disaster preparedness.

Chapter 6:
Extreme Value Modeling with GPD

Understanding and modeling extreme events is vital across various disciplines—from finance to environmental science. Among the suite of statistical tools designed for this purpose, the GPD plays a pivotal role in characterizing the behavior of values that exceed a high threshold. This chapter explores the power and flexibility of GPD through two complementary approaches: a controlled simulation study and a real-world application on daily rainfall data from Meghalaya, India. Together, these analyses underscore the practical relevance of GPD in modeling tail behavior and support its use in risk assessment and extreme value forecasting.

6.1 GPD & GEVD Fitting through Simulated Data

This simulation study aims to evaluate the performance of the GPD and GEVD in modeling extreme events. We generate synthetic data to mimic rare events and fit both the GPD and GEVD models to estimate their parameters.

The goal is to assess how well these distributions capture tail behavior and the accuracy of their parameter estimates. Model fit will be evaluated using selection criteria like the AIC, providing insights into the best-fitting model for the simulated data. This study sets the foundation for applying these models to real-world datasets.

In this analysis, we used the approach outlined in the Generalized Pareto and GEV Distributions – MATLAB Documentation for simulating data from the GPD and GEVD . Although we followed the methodology described in the MATLAB guide, we implemented the analysis in R. Specifically, we used R's equivalent functions to simulate the data and estimate parameters via MLE for both GPD and GEVD.

Algorithm for Modeling Extremes using GEVD and GPD Approaches

1. **Generate Data**: Generate 1000 observations from a *Student's t-distribution* with 3 degrees of freedom.

2. **Divide into Blocks**: Divide the generated data into *50 blocks*, each of size 20, for use in the GEVD model.

3. **Extract Maxima or Exceedances**:

    - For GPD (POT), set the threshold at the *95th percentile* and identify all values that exceed the threshold. These are the *exceedances*.

    - For GEVD (Block Maxima), extract the *maximum* value from each block.

4. **Fit Models and Compute AIC/BIC**: Fit a GPD model to the excesses, fit a GEVD model to the block maxima, and calculate the AIC and BIC for both models.

R code

```
# ================================================
# STEP 1: SIMULATE DATA
# ================================================
# Simulate heavy-tailed data
n_obs <- 1000
baseline_data <- rt(n_obs, df = 3)  # Heavy-tailed (df=3 for fat
   tails)
# (A) For GEV: Take block maxima
n_blocks <- 50  # 50 years of data
block_size <- n_obs / n_blocks  # Observations per block
block_maxima <- sapply(1:n_blocks, function(i) {
  max(baseline_data[((i-1)*block_size + 1):(i*block_size)])})
# (B) For GPD: Take exceedances over a high threshold
threshold <- quantile(baseline_data, 0.95)  # 95th percentile
exceedances <- baseline_data[baseline_data > threshold] - threshold
# ================================================
# STEP 2: FIT GEV MODEL
# ================================================
# Fit GEV using MLE
gev_fit <- fevd(block_maxima, type = "GEV")
# Get summary and AIC
print(summary(gev_fit))
gev_aic <- gev_fit$results$AIC
# ================================================
# STEP 3: FIT GPD MODEL
# ================================================
# Fit GPD using MLE
gpd_fit <- fevd(exceedances, threshold = 0, type = "GP")
# Get summary and AIC
print(summary(gpd_fit))
gpd_aic <- gpd_fit$results$AIC
# ================================================
# STEP 4: COMPARE MODELS
# ================================================
# Compare AIC values
aic_comparison <- data.frame(
  Model = c("GEV (Block Maxima)", "GPD (Threshold Exceedances)"),
  AIC = c(gev_aic, gpd_aic)
)
cat("\nModel Comparison:\n")
print(aic_comparison)
# ================================================
# STEP 5: VISUALIZATION
# ================================================
# GEV vs Empirical (Block Maxima)
```

```
gev_x <- seq(min(block_maxima), max(block_maxima), length.out = 100)
gev_cdf <- pevd(gev_x,
                loc = gev_fit$results$par[1],
                scale = gev_fit$results$par[2],
                shape = gev_fit$results$par[3],
                type = "GEV")
empirical_cdf_gev <- ecdf(block_maxima)(gev_x)
# GPD vs Empirical (Exceedances)
gpd_x <- seq(0, max(exceedances), length.out = 100)
gpd_cdf <- pevd(gpd_x,
                scale = gpd_fit$results$par[1],
                shape = gpd_fit$results$par[2],
                type = "GP")
empirical_cdf_gpd <- ecdf(exceedances)(gpd_x)
# Plot GEV Fit
p1 <- ggplot(data.frame(x = gev_x, GEV = gev_cdf, Empirical =
  empirical_cdf_gev), aes(x = x)) +
  geom_line(aes(y = GEV, color = "GEV Fit"), linewidth = 1) +
  geom_line(aes(y = Empirical, color = "Empirical"), linewidth = 1)
     +
  labs(title = "GEV Fit vs Empirical CDF (Block Maxima)",
       x = "Value", y = "CDF") +
  scale_color_manual(values = c("GEV Fit" = "red", "Empirical" = "
     blue")) +
  theme_minimal() +
  theme(legend.position = "top")
# Plot GPD Fit
p2 <- ggplot(data.frame(x = gpd_x, GPD = gpd_cdf, Empirical =
   empirical_cdf_gpd), aes(x = x)) +
  geom_line(aes(y = GPD, color = "GPD Fit"), linewidth = 1) +
  geom_line(aes(y = Empirical, color = "Empirical"), linewidth = 1)
     +
  labs(title = "GPD Fit vs Empirical CDF (Threshold Exceedances)",
       x = "Exceedance", y = "CDF") +
  scale_color_manual(values = c("GPD Fit" = "red", "Empirical" = "
     blue")) +
  theme_minimal() +
  theme(legend.position = "top")
```

Model Comparison: AIC-based Evaluation

Table 6.1: Comparison of GEV and GPD models based on AIC

Model	AIC Value	Interpretation
GEV (Block Maxima)	200.4708	Higher AIC – less preferred
GPD (Threshold Exceedance)	146.7855	Lower AIC – better model fit

The AIC helps in model selection by penalizing model complexity. Since the GPD model has a lower AIC value, it provides a better trade-off between goodness-of-fit and model simplicity.

Hence, the GPD model is preferred for this dataset.

Model Evaluation: Based on CDF Fit

The graphical comparison between the empirical and theoretical CDFs for both the GEVD and GPD models provides valuable insight into their goodness of fit. In Figure 6.1, the GEVD model is fitted to block maxima data using the block maxima method. The plot shows that the theoretical CDF closely follows the empirical CDF, especially in the upper tail, indicating that GEVD is able to capture the behavior of extremes fairly well. Similarly, Figure 6.2 shows the GPD model fit using the POT approach. The theoretical and empirical CDFs exhibit a strong agreement, suggesting that GPD also fits the exceedance data well.

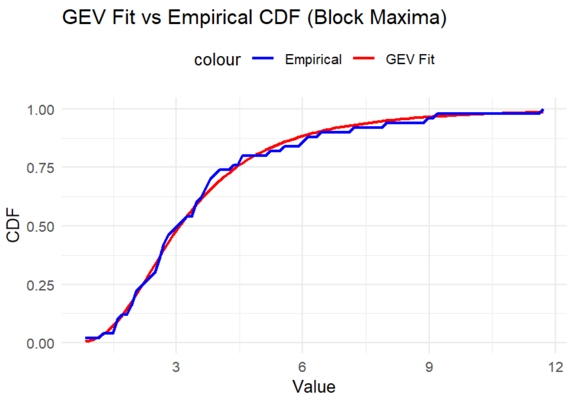

Figure 6.1: Empirical vs Theoretical CDF for GEVD

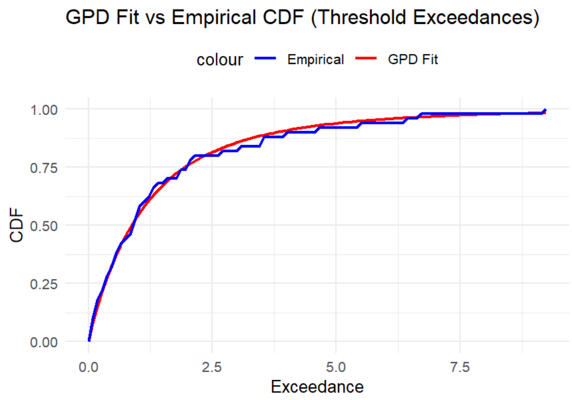

Figure 6.2: Empirical vs Theoretical CDF for GPD

Additionally, the comparison of AIC values reveals that the GPD model yields a lower AIC than the GEVD model, indicating a better trade-off between model fit and complexity. Therefore, while both models show good visual fits, the GPD model is statistically preferred for modeling the extreme values in this dataset. This supports the use of the POT approach with GPD when threshold exceedances are of primary interest in heavy-tailed environmental data.

6.2 Modeling Rainfall Exceedances in Meghalaya Using GPD

Extreme rainfall events present significant challenges, especially in regions with monsoon variability, leading to socio-economic impacts such as flooding, infrastructure damage, and agricultural disruptions. Understanding and modeling these extremes is essential for disaster preparedness and climate resilience.

In this section, we apply the GPD to model daily rainfall data, focusing on values exceeding a selected threshold. The GPD effectively captures the tail behavior of extreme values, crucial for understanding the frequency and severity of rare but impactful rainfall events. By modeling exceedances, we isolate extreme rainfall events while excluding smaller, less relevant values.

The analysis starts with estimating the GPD parameters by fitting the model to the exceedances data, followed by diagnostics to assess model fit. We also compare the GPD with the GEVD using the AIC for model selection, identifying the best model for extreme rainfall events in the region.

This application highlights how statistical extreme value theory, particularly the GPD, provides insights into environmental extremes. Modeling extreme rainfall events enhances our understanding of their behavior and aids in planning for their impacts, supporting climate resilience efforts and improving disaster preparedness strategies.

In their study on Tanzania, Iyamuremye et al. [2019] utilized the GPD model for extreme rainfall analysis. This methodology is similarly adopted here to analyze rainfall extremes in Meghalaya.

Daily Rainfall Data for Meghalaya (2019–2024)

Table 6.2: Selected Daily Rainfall Records from 2019 to 2024

Year	Month	Day	Rainfall (mm)
2019	1	1	0.00
2019	1	2	0.00
⋮	⋮	⋮	⋮
2020	9	22	89.59
2020	9	23	81.36
2020	9	24	103.21
2020	9	25	109.51
2020	9	26	64.67
2020	9	27	25.07
⋮	⋮	⋮	⋮
2024	12	30	0.00
2024	12	31	0.00

Note: Only a subset of the full dataset is shown for illustration. The complete dataset spans six years of daily rainfall observations.

Description: The dataset presents daily rainfall (in millimeters) recorded at Cherrapunji station from January 1, 2019, to December 31, 2024. Each entry includes the year, month, day, and corresponding rainfall value. Most entries have zero rainfall, indicating dry days, with few days recording measurable precipitation. This dataset is used to analyze extreme rainfall events and estimate return levels using extreme value theory models such as GPD and GEV.

The primary purpose of analyzing this dataset is to understand the behavior of extreme rainfall events and to estimate return levels that can inform infrastructure design, disaster preparedness, and climate resilience planning in the region.

Algorithm: GPD Modeling and Return Level Estimation for Rainfall Data

1. **Load Required Packages:** Load necessary R libraries such as `extRemes` for GPD modeling, `tidyverse` for data manipulation, `lubridate` for handling dates, and `readxl` to import Excel data.

2. **Read and Preprocess Data:**
   - Import daily rainfall data using `read_xlsx()`.
   - Construct a proper `Date` column using `make_date()` from the Year, Month, and Day columns.
   - Sort the data chronologically.
   - Check and report missing values.
   - Summarize the rainfall variable using descriptive statistics.

3. **Visualize Rainfall Data:**
   - Plot a time series graph of daily rainfall over time.
   - Plot a histogram to understand the distribution of daily rainfall.

4. **Set Threshold for POT Method:** Define a high threshold (e.g., the 95th percentile of rainfall values) for POT modeling.

5. **Fit GPD Model:** Fit a GPD to the threshold exceedances using `fevd()` with type = ”GP”, specifying time units and return period basis.

6. **Summarize Fitted Model:**
   - Extract parameter estimates (scale and shape).
   - Obtain standard errors, confidence intervals, and model diagnostics.

7. **Estimate Return Levels:**
   - Choose return periods (e.g., 2, 5, 10, 20, 50, 100 years).
   - Compute return levels using `return.level()`.
   - Plot return levels against return periods to interpret risk levels.

8. **Confidence Intervals:**
   - Obtain confidence intervals for GPD parameters using `ci()`.
   - Compute confidence intervals for return levels to understand uncertainty.

R code:

```
# Load necessary libraries
library(extRemes)  # For extreme value analysis
library(tidyverse) # For data manipulation
library(lubridate) # For date handling
library(readxl)
# Read the data
rainfall_data <- read_xlsx("D:/MSc/#MSc 2nd year/ST-525-Research
   project/Data/corrected/Meghalaya.xlsx")
# For this example, create a dataframe from the sample data shown
# Create a Date column
rainfall_data <- rainfall_data %>%
  mutate(Date = make_date(Year, Month, Day))
# Sort by date
rainfall_data <- rainfall_data %>% arrange(Date)
# Check for missing values
sum(is.na(rainfall_data$Rainfall))
# Basic summary statistics
summary(rainfall_data$Rainfall)
# Plot the time series
ggplot(rainfall_data, aes(x = Date, y = Rainfall)) +
  geom_line() +
  labs(title = "Daily Rainfall in Meghalaya", x = "Date", y = "
     Rainfall (mm)") +
  theme_minimal()
# Histogram of rainfall amounts
ggplot(rainfall_data, aes(x = Rainfall)) +
  geom_histogram(bins = 50, fill = "blue", alpha = 0.7) +
  labs(title = "Distribution of Daily Rainfall", x = "Rainfall (mm)"
     , y = "Frequency") +
  theme_minimal()
# Fit GPD to exceedances over threshold
threshold <- quantile(rainfall_data$Rainfall,0.95)
gpd_fit <- fevd(rainfall_data$Rainfall, threshold = threshold, type
   = "GP",
                time.units = "days", period.basis = "year")
# Summary of the fitted model
summary(gpd_fit)
# Calculate return levels
return_periods <- c(2, 5, 10, 20, 50, 100)  # In years
return_levels <- return.level(gpd_fit, return.period = return_
   periods)
# Plot return levels
plot(return_periods,return_levels,type = "b", pch = 19, col = "
   steelblue",
     xlab = "Return Period (Years)",
     ylab = "Return Level (mm)",
     main = "Return Level Plot")
grid()
# Confidence intervals for parameters
```

```
ci(gpd_fit, type = "parameter")
# Confidence intervals for return levels
ci(gpd_fit, return.period = return_periods)
```

Results

Table 6.3: Statistical Summary of Daily Rainfall in Meghalaya

Variable	Period	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Std
Rainfall	2019-2024	0.00	0.00	2.19	11.58	13.78	252.72	21.82

The statistical summary of daily rainfall in Meghalaya (2019-2024) indicates considerable variability, with a mean of 11.58 mm and a maximum of 252.72 mm. Most days had light rainfall, with a median of 2.19 mm, but extreme events were occasional, as reflected by the high standard deviation of 21.82 mm.

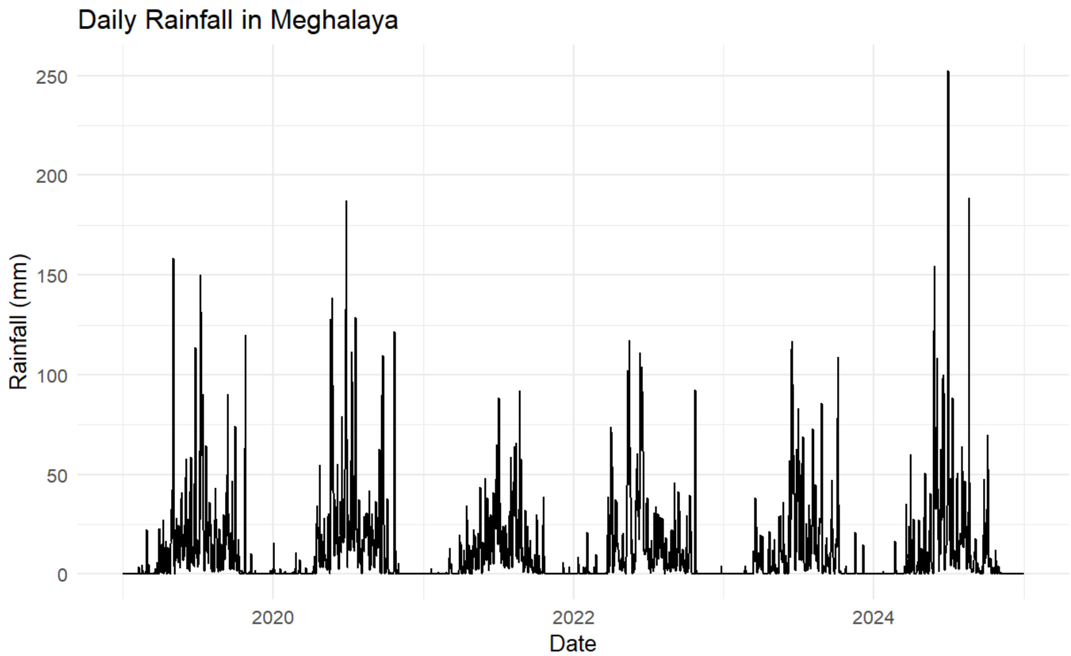

Figure 6.3: Time Series Plot of Daily Rainfall in Meghalaya (2019–2024)

The time series plot of daily rainfall in Meghalaya (2019 - 2024) shows significant variability in rainfall patterns over the years. Several periods exhibit heavy rainfall events, indicated by large spikes in the plot, especially in 2020 and 2024. These peaks suggest occasional extreme rainfall days, while the rest of the time, the rainfall amounts are relatively lower and more consistent. This pattern is indicative of sporadic heavy rainfalls, which are typical in regions with a monsoon climate.

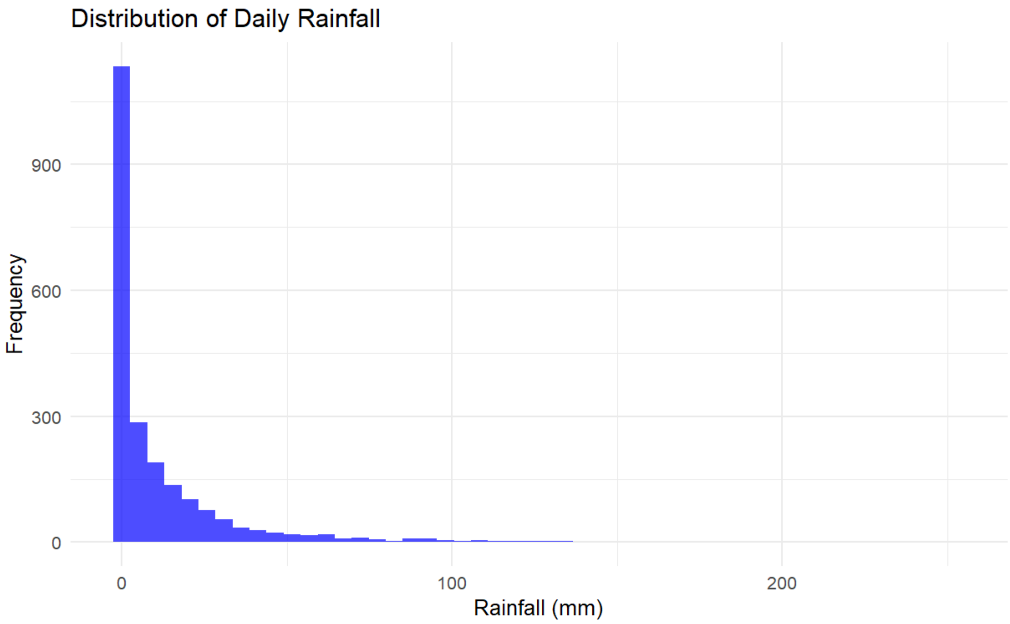

Figure 6.4: Distribution of Daily Rainfall in Meghalaya (2019–2024)

The histogram shows the distribution of daily rainfall in Meghalaya from 2019 to 2024. Most of the rainfall values are concentrated around low amounts, with a large frequency of days receiving very little rainfall. As the rainfall amount increases, the frequency decreases, suggesting that heavy rainfall events are rare. The distribution appears to be right-skewed, indicating that extreme rainfall events, though less frequent, are significant when they occur.

For modeling extreme rainfall events, we selected the threshold for the GPD as the 95th percentile of the rainfall data. This choice ensures that only the most significant rainfall excesses are considered, focusing on the tail of the distribution to capture extreme events accurately.

Table 6.4: Statistical Summary of Rainfall Excesses in Meghalaya

Variable	Period	Min	1st Qu.	Median	Mean	3rd Qu.	Max	SD
Rainfall Excesses	2019-2024	54.93	62.58	76.41	86.52	99.50	252.72	32.47

The rainfall excesses in Meghalaya from 2019 to 2024 ranged from 54.93 mm to 252.72 mm, with a mean of 86.52 mm and a standard deviation of 32.47 mm, indicating moderate variability.

Table 6.5: Comparison of GPD and GEV

Parameter	GPD (Estimated)	GEV (Estimated)
Location	1.399861	1.399861
Scale	31.246037	7.833779
Shape	0.017044	5.595924
AIC	985.0968	7760.504
BIC	990.4978	7777.582

The comparison of the GPD and GEVD models shows that GPD has a significantly lower AIC and BIC compared to GEV. These lower values for GPD indicate a better fit to the data. Therefore, based on these criteria, GPD performs better than GEV in this scenario. Consequently, further analysis will focus on using the GPD for return level calculations.

Table 6.6: 95% CI for Return Levels and Estimates

Return Period	**Estimated Return Level (mm)**	**95% CI**
2-year	170.7728	(141.3740, 200.1717)
5-year	201.4548	(154.9116, 247.9980)
10-year	224.9852	(161.3930, 288.5775)
20-year	248.7953	(164.6178, 332.9728)
50-year	280.7053	(163.8285, 397.5822)
100-year	305.1776	(159.3476, 451.0076)

The 2-year return level of 170.77 mm indicates that rainfall exceeding this amount has a 50% chance of occurring in any given year, while the 10-year return level of 224.99 mm suggests that rainfall exceeding this amount has a 10% chance of occurring annually. As the return period increases, the expected rainfall amount becomes higher, indicating more extreme events with lower probability. These return levels are important for understanding and preparing for extreme rainfall events over different time spans.

Figure 6.5: Return Level Plot for Rainfall Data in Meghalaya

The Return Level Plot shows that as the return period increases, the estimated rainfall level also rises, indicating a higher intensity of rare rainfall events. The curve suggests a nonlinear increase, with extreme values becoming more pronounced at longer return periods. This trend highlights the growing severity of rainfall events over time.

Summary

- Understanding the Foundations of Extreme Events:Initiated the study by defining extreme events and introduced the statistical concept of Return Period, laying the theoretical groundwork for identifying and analyzing rare environmental occurrences. My work focused on collecting rainfall data from multiple Indian regions and performing a preliminary analysis of maximum rainfall patterns to estimate the likelihood of future events.

- Classification and Documentation of Major Extreme Events : Categorized various types of extreme weather phenomena (floods, droughts, cyclones, etc.) and identified event-specific thresholds. Compiled case studies of significant Indian and global disasters, contributing a comparative narrative that connects climate trends with historical impact.

- Assessing Risks and Impacts Under Climate Change:Analyzed death rates and economic burdens caused by extreme weather events using statistical indicators such as per million mortality rates. Developed insights into how vulnerability and exposure intersect with hazard intensity. Highlighted the role of statistical tools in informing adaptive public health and disaster policies.

- Modeling with Extreme Value Distributions: Applied theoretical EVT models including Gumbel, Fréchet, and Weibull distributions to analyze environmental data. Focused on parameter interpretation and used graphical comparisons to understand the tail behavior of different distribution types.

- Using Generalized Pareto Distribution & POT Approach: Conducted detailed statistical modeling using the Peaks Over Threshold (POT) approach. Estimated return levels and parameters via Maximum Likelihood Estimation (MLE) and compared GEVD and GPD methods. Utilized R packages like `ismev`, `evir`, and `extRemes` for practical implementation, contributing key computational results to the study.

- Simulation and Real Data Application:Demonstrated EVT applications through simulations and a case study on Meghalaya's rainfall data. Modeled exceedances and derived return levels for future risk estimation. Evaluated the model fit and validated findings with confidence intervals, contributing original insights based on real environmental data.

Bibliography

C. S. Bahinipati and A. K. Gupta. Methodological challenges in assessing loss and damage from climate-related extreme events and slow onset disasters: Evidence from india. *International Journal of Disaster Risk Reduction*, 83:103418, 2022.

S. G. Coles and J. A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 53(2):377–392, 1991.

S. G. Coles and D. Walshaw. Directional modelling of extreme wind speeds. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 43(1):139–157, 1994.

R. Davis and S. Resnick. Extremes of moving averages of random variables from the domain of attraction of the double exponential distribution. *Stochastic processes and their applications*, 30(1):41–68, 1988.

M. Fischedick, J. Roy, A. Acquaye, J. Allwood, J.-P. Ceron, Y. Geng, H. Kheshgi, A. Lanza, D. Perczyk, L. Price, et al. Industry in: Climate change 2014: Mitigation of climate change. contribution of working group iii to the fifth assessment report of the intergovernmental panel on climate change. technical report. *Environmental Science & Policy*, 2014.

J. Galambos. The development of the mathematical theory of extremes in the past half century. *Theory of Probability & Its Applications*, 39(2):234–248, 1995.

I. M. Goklany. Death and death rates due to extreme weather events. *Civil Society Report on Climate Change*, 47, 2007.

E. J. Gumbel. *Statistics of extremes*. Columbia university press, 1958.

E. Iyamuremye, J. Mung'atu, P. Mwita, et al. Extreme value modelling of rainfall using poisson-generalized pareto distribution: A case study tanzania. 2019.

H. Joe, R. L. Smith, and I. Weissman. Bivariate threshold methods for extremes. *Journal of the royal statistical society: series B (methodological)*, 54(1):171–183, 1992.

S. Kotz and S. Nadarajah. *Extreme value distributions: theory and applications*. world scientific, 2000.

V. Murray and K. L. Ebi. Ipcc special report on managing the risks of extreme events and disasters to advance climate change adaptation (srex), 2012.

W.-H. Nam, J.-Y. Choi, S.-H. Yoo, and M.-W. Jang. A decision support system for agricultural drought management using risk assessment. *Paddy and Water Environment*, 10(3):197–207, 2012.

J. PICKANDs III. A method is presented for making statistical inferences about the upper tail of a distribution function. it is useful for estimating the probabilities of future extremely large observations. the method is applicable if the underlying distribution function satisfies a condition which holds for all. *The Annals of Statistics*, 3(1):119–131, 1975.

M. Pinheiro and R. Grotjahn. An introduction to extreme value statistics. *Tutorial at University of California at Davis*, 2015.

M. Rummukainen. Changes in climate and weather extremes in the 21st century. *Wiley Interdisciplinary Reviews: Climate Change*, 3(2):115–129, 2012.

J. A. Tawn. Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415, 1988.

X.-C. Yuan, X. Sun, U. Lall, Z.-F. Mi, J. He, and Y.-M. Wei. China's socioeconomic risk from extreme events in a changing climate: a hierarchical bayesian model. *Climatic Change*, 139: 169–181, 2016.

X.-C. Yuan, Y.-M. Wei, B. Wang, and Z. Mi. Risk management of extreme events under climate change. *Journal of Cleaner Production*, 166:1169–1174, 2017.