London's Global University

IRDR0014 MRes Independent Research Project

Integrated Multi-Hazard and Vulnerability Modelling for Flood Risk Assessment in the US Gulf Coast

Student number: 23218265

Candidate number: JBGV2

Academic Supervisor: Prof Peter Sammonds, Dr Shipra Jain

Dissertation submitted to the Department of Risk and Disaster Reduction in partial fulfilment of the requirements for the Master of Research

University College London (UCL)

Word count: 17,079

DECLARATION

I declare that the following work is my own and, where the work of others has been used, it has been clearly identified.

I have received approval for *ethics exemption*.

Abstract

This thesis proposes a multi hazard modelling framework to address the increasing risk of flood-related damages in the US Gulf Coast region due to climate change, urbanisation, and extreme weather events. This research develops a statistical per event-based catastrophe model on predicting the financial losses from flooding triggered by the heavy precipitation/rainfall from abnormal (high) windspeed via integrating statistical (Extreme Value Theory and Copula analysis) and machine learning methodologies (Generalised Linear Model). This study focuses on the US Gulf Coast, particularly the states of Texas, Louisiana and Mississippi. The proposed assessment is structured into two distinct yet complementary objectives: the hazard and vulnerability models. The hazard model aims to calculate the annual joint occurrence probability for wind-driven floods by capturing the complexities and dependencies of how floods will be driven and impacted by windspeed. The vulnerability model aims to predict potential losses for single buildings or a portfolio of properties under different building characteristics, and it is evaluated using cross-validation methods to examine its reliability and accuracy. The outcome is the financial losses from the specific intensity of natural disaster events and the exceedance probability for specific thresholds. The outcome shows the risk of extreme windspeed events across three states on the US Gulf Coast. It provides valuable insights for understanding, predicting, and managing risks associated with wind-related hazard and their cascading effect, contributing to urban planning, disaster management, and the insurance and reinsurance industry. Also, this research contributes to developing a multi-hazard flood risk assessment framework and offers a solid foundation of non-commercial tools for researchers and policymakers to better manage coastal flood risks.

Table of contents

List of Figures

List of Tables

Acknowledgements

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude to my parents, Gang Liu and Donghong Luo, for their continued support, encouragement, and understanding throughout my academic journey. Their belief in me has been my greatest source of strength.

I am profoundly grateful to my thesis supervisors, Prof Peter Sammonds and Dr Shipra Jain, for their invaluable guidance, insightful feedback, and constant encouragement. Their expertise and mentorship have been instrumental in completing this thesis.

I would also like to thank my course and research module leader, Dr Sarah Dryhurst, for her invaluable guidance on administrative matters, and for the time she took to answer my questions regarding the submission process.

Lastly, I wish to thank my partner Wanchen Shi (Odile), for her encouragement, emotional support and for making this journey both enjoyable and memorable.

Furthermore, the code and model used in this thesis are available on GitHub at https://github.com/AAn-Liu/Multi.git. I hope this will be a helpful resource for others interested in this area of research.

1. Introduction

1.1. Background

The rapid increase in global population and the ongoing trend of urbanisation are leading to a higher number of people being exposed to flood risks. As of November 2022, the global population exceeded 8 billion and continues to grow (United Nations Department of Economic and Social Affairs, Population Division, 2022 By 2050, it is expected that over two-thirds of the world's population will live in urban areas, significantly raising the number of individuals vulnerable to urban flooding (Gu, 2019; Tellman et al., 2021). The impacts of climate change are intensifying these flood risks by increasing the frequency and severity of extreme weather events, driven by an unprecedented rate of global warming due to human activities. Each additional 0.5°C of warming results in notably higher intensity and frequency of such extreme events. The Intergovernmental Panel on Climate Change (IPCC) reports that since the early 20th century, the global average sea level has risen by approximately 20 centimetres, with projections suggesting a possible rise of up to 1 meter by 2100 under high-emission scenarios (IPCC, 2021). These changes pose significant threats to regions such as the Gulf Coast, where even minor increases in sea level can lead to substantial increases in both the intensity and possibility of floods.

Additionally, climate change is expected to intensify the hydrological cycle, resulting in more extreme precipitation events. This trend is evident in the increasing incidence of heavy rainfall and flash flooding on the Gulf Coast. Studies indicate that the intensity of extreme precipitation events in the southeastern United States has increased by approximately 27% over the last 50 years, largely due to warmer air temperatures that enhance the moisture-holding capacity of the atmosphere (US Global Change Research Program, 2017).

Environmental degradation, including the loss of wetlands and deforestation, further aggravates flood risks in the Gulf Coast. Wetlands act as natural buffers, absorbing excess rainfall and reducing the impact of storm surges. However, the Gulf Coast has lost substantial wetlands due to urban development, industrial activities, and rising sea levels. According to the US Geological Survey, the region loses an average of 80 square

kilometres of wetlands each year, diminishing its natural flood mitigation capacity (US Geological Survey, 2020). In addition, deforestation and land degradation in the Gulf Coast's watershed areas contribute to increased runoff and sedimentation in rivers, leading to reduced channel capacity and higher flood risks. Effective flood risk management in the Gulf Coast thus requires integrated approaches that address both environmental conservation and sustainable urban development.

In the US, 41 million people were exposed to floodplain risks in 2018, and this number could grow to 60 million by 2050 due to population growth and climate change (Wing et al., 2018). The economic costs of urban flooding are projected to increase dramatically in the coming decades, with annual global losses from coastal urban flooding that could reach $1 trillion by 2050 (Hallegatte et al., 2013).

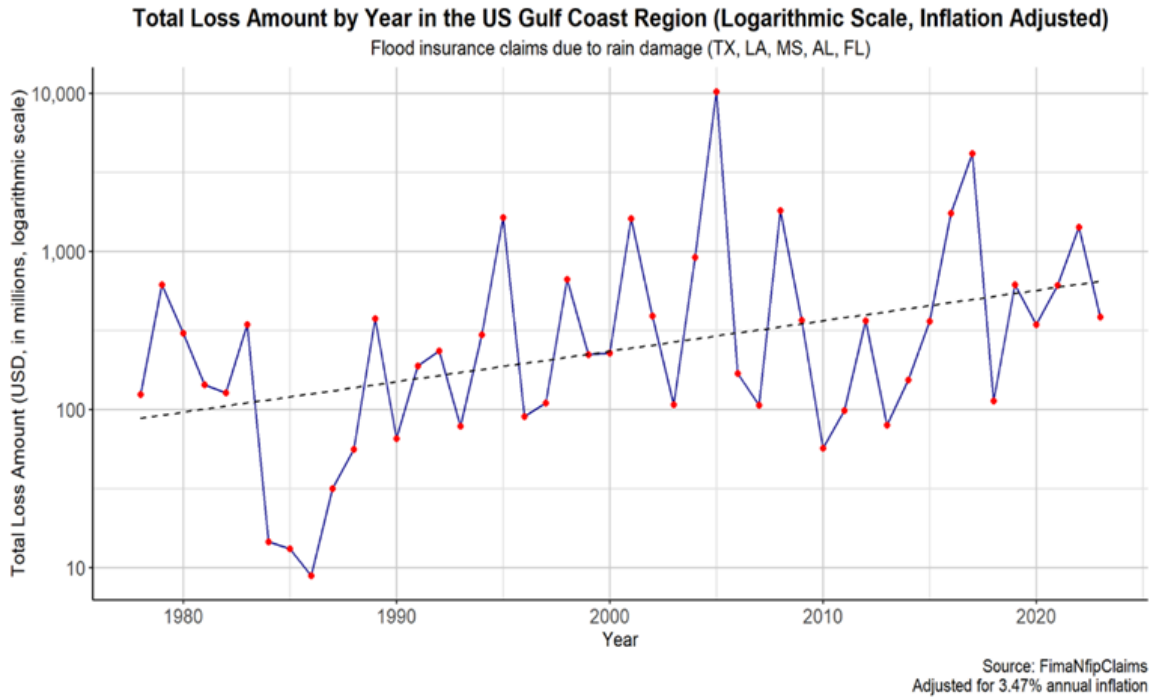

Figure 1.1: Total Loss Amount by Year in the US Gulf Coast Region

Figure 1.1 highlights the temporal trends in flood insurance claims due to rain damage in the US Gulf Coast states. The logarithmic scale on the y-axis shows significant variability in annual losses, with notable spikes in certain years indicating major flood events such as Tropical Storm Allison, 2001; Hurricane Katrina & Hurricane Rita, 2005; Hurricane Harvey,

2017. The dashed trend line indicates an overall increasing trend in the total loss amount, reflecting either an increase in the frequency or severity of flood events or changes in the insured value of properties. By adjusting for inflation, Figure 2.1 provides a clearer picture of the economic impact of flood losses over time, allowing for better comparison across different years.

1.2. Problem Statement

The devastation caused by Hurricane Harvey in 2017 resulted in unprecedented rainfall and flooding in Texas and Louisiana, underscores the increasing threat of such extreme weather events (Blake & Zelinsky, 2018). Over the past decade, the Gulf Coast has witnessed a compound annual growth rate of approximately 4% in exposure, and this growth rate is expected to increase the likelihood of both the intensity and frequency of severe hurricanes (Categories 4 and 5 on the Saffir-Simpson Scale) in the future (Sousounis & Little, 2017). According to recent NOAA predictions, the 2024 Atlantic hurricane season is expected to be above normal due to near-record warm ocean temperatures and La Niña conditions, with 17 to 25 named storms, 8 to 13 becoming hurricanes, and 4 to 7 major hurricanes (NOAA, 2024). The Gulf Coast's unique geographical and meteorological conditions, influenced significantly by the Gulf of Mexico, necessitate a focused study on predicting and mitigating flood losses driven by wind and rainfall (Keim & Muller, 2009).

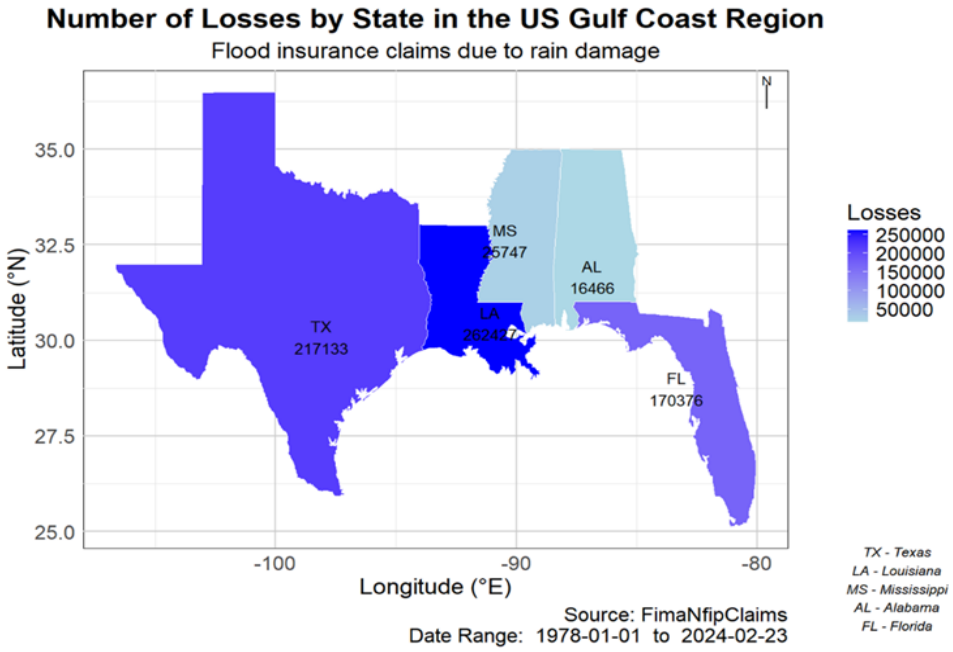

Figure 1.2:Number of Losses by State in the US Gulf Coast Region

Figure 1.2 highlights the spatial distribution of flood insurance claims due to rain damage across the US Gulf Coast states between January 1978 to February 2024. The colour gradient indicates the frequency of losses, with darker shades representing higher numbers of claims. In these states, Louisiana (LA) has the highest number of claims (262,427), followed by Texas (TX) with 217,133 claims. The figures for other states are as follows: Florida (FL) with 170,376 claims, Mississippi (MS) with 25,747 claims, and Alabama (AL) with 16,466 claims.

Insurance and reinsurance play crucial roles in managing and mitigating flood risks by providing financial protection and reducing premiums for risk reduction measures (Kunreuther & Michel-Kerjan, 2009; Surminski & Oramas-Dorta, 2014). However, the effectiveness of these financial instruments depends on the accuracy of flood risk assessments and the integration of comprehensive vulnerability data (Botzen et al., 2009). There is a need for improved flood insurance schemes that account for the changing risk landscape due to climate change and urban development (Lamond & Penning-Rowsell, 2014).

Furthermore, while there is a variety body of research on physical flood hazards, there is less emphasis on the vulnerability of specific populations and infrastructure. Vulnerability assessments often lack the granularity needed to accurately identify at-risk buildings and communities. This deficiency hampers the development of targeted mitigation and adaptation strategies that could significantly reduce the impacts of flooding (Mechler et al., 2014).

Despite the catastrophe models used by the insurance industry, which are still not perfect, they are improving and have been used to manage trillions of euros in coverage globally (Michel-Kerjan, 2012). According to de Ruig et al. (2022), utilising these models typically requires substantial annual fees, which can be prohibitive for many stakeholders, including smaller municipalities and independent researchers. Additionally, the technical complexity of these models poses challenges for those without specialised expertise in flood risk modelling, limiting their accessibility and practical utility. Mechler and Bouwer (2015) argue that adequate funding is essential for implementing comprehensive disaster risk reduction measures, which can significantly reduce vulnerability and enhance resilience.

1.3. Research Aim and Objectives

To address these challenges, this research seeks to create a non-commercial financial flood loss prediction model for researchers, focusing on two main objectives: firstly, developing a hazard model to quantify the impact of windstorms and heavy rainfall, and secondly, constructing a machine learning-based vulnerability model for precise loss estimation at the level of individual buildings. The development of a non-commercial model for researchers to assess resilient strategies for managing coastal risks has become increasingly vital. The study areas examined in this paper are Texas, Louisiana, and Mississippi.

1.4. Thesis Structure

The structure of this study is as follows: Section 2 provides a review of existing literature to identify gaps in multi-hazard flood modelling. Section 3 outlines the methodology, detailing the data sources used for model evaluation, dataset visualisation, and model development techniques. Section 4 presents the results of the model. Section 5 discusses the implications of our findings for urban planning and disaster management, highlights potential limitations of the model, and offers directions for future research and recommendations. Finally, Section 6 is the conclusion.

2. Literature Review

2.1. Climate Change and Flood Predictions

Scaife et al. (2017) indicate that the predictability of storminess is at the seasonal scale, from a month to three months. However, winter wind speed and rainfall are trending to increase in the second half of the 21st century due to projections of increased frequency of storm activities (UKCP18, Lowe et al., 2018). Furthermore, Fung et al. (2018), go on to point out that the rainfall pattern would be seasonal, which means the rainfall would be more in winter and less in summer; however, in general, there is a trend of increase in intensity and frequency of extreme rainfall events (Met Office, 2018; 2019a). According to McDonald (2011), future models (CMIP6) should take up a multi-model scenario to bring down uncertainties with regard to the prediction of storms under climatic change. These predictions are expected because global warming will induce changes in the hydrological cycle, with floods and extreme precipitation likely to increase (Tabari, H. 2020; Madsen et al., 2014).

Recent research indicates the growing frequency and intensity of natural hazards due to climate change and urbanisation, especially in highly populated areas. The traditional single-hazard model is less able to deal with the complex coupling of multiple hazards, such as the co-occurrence of floods, landslides, and storm surges. Gallina et al. (2016) emphasise that multi-hazard risk assessment is critically necessary, pointing to significant gaps in existing methodologies that often do not consider the cumulative and cascading effects of multiple hazards. For example, Johnson et al. (2016) illustrate the challenges of assessing multi-hazard risks in urban areas, noting that different hazards, like heat waves, typhoons, and landslides, may interact in complex ways, particularly in socio-economically diverse districts like those in Hong Kong. Similarly, Kappes et al. (2012) discuss how the interconnected nature of urban vulnerabilities, such as those driven by socioeconomic disparities, exacerbates the impacts of overlapping hazards, necessitating integrated approaches that consider both the spatial and temporal dimensions of risk. Collectively, this body of work identifies a critical research gap: the lack of comprehensive models that can integrate multiple hazards under an adaptive framework, particularly in the context of

climate change. Addressing this gap is essential for developing more effective risk mitigation strategies in urban areas.

The increasing complexity of natural hazards due to climate change necessitates the development of multi-hazard models that can account for the interdependencies between various risks. Traditional risk assessment approaches, such as Probabilistic Risk Assessment (PRA) and Deterministic Risk Assessment (DRA), have been foundational in evaluating individual hazards, but they fall short when applied to interconnected and cascading risks. Recent literature highlights the shift towards integrated frameworks that combine these approaches within a multi-hazard context. For instance, Jongman et al. (2014) propose a multi-risk assessment framework, emphasising the importance of considering the dependencies across different hazards and socioeconomic factors. This framework advocates for a system-of-systems approach, where risks are evaluated not only individually but also in relation to their potential to trigger cascading failures across sectors and regions. Additionally, Johnson et al. (2016) discuss the application of PRA and DRA in urban environments, where the interaction between multiple hazards, such as flooding and landslides, requires a more dynamic and comprehensive assessment methodology. These frameworks guide the development of multi-hazard models that are better equipped to manage the complexities of contemporary risk landscapes, offering a more holistic approach to disaster risk reduction.

2.2. Multivariate Flood Hazard Modelling

Eilander et al. (2023) present a framework for compound flood hazard modelling that has global applicability, emphasising the need to consider multiple drivers of flooding, such as extreme rainfall, river discharge, and coastal water levels. This study underscores the importance of understanding interactions between different flood drivers for effective risk management. Wahl et al. (2015) discuss the combined effects of sea-level rise, storm surges, and extreme precipitation on flood hazards, highlighting the need for comprehensive flood risk models that incorporate multiple interacting drivers.

In terms of multivariate flood hazard modelling, two major categories are Dynamical Models and Statistical Models. Dynamical models focus on detailed physical simulations and capture the complexity of the physical environment. These models require high-quality data, proper model setup, and parameterisation to ensure reliable dynamic flood simulations (Nkwunonwo et al., 2020). On the other hand, statistical models aim to understand and quantify the relationship between flood hazards and different drivers, using physical data to produce flood maps or non-physical data, such as rainfall and sea levels, to estimate occurrence probabilities (Khosravi et al., 2016; Towe et al., 2016).

The Weather Research and Forecasting (WRF) model is one of the most commonly utilised dynamical models for rainfall flood forecasting in both research and practical applications (Jiyang et al., 2017). However, while simulating winds, surges and extreme rainfall using the WRF model from Hurricane Isabel (2003), the simulations failed to capture the outer rainbands accurately, which were critical in contributing to wind damage and urban flooding (Lin et al., 2010). Furthermore, the WRF model underestimates 50% to 60% of simulated rainfall events (Moya-Álvarez et al., 2018). To address this issue, Couasnon et al. (2020) explore the statistical dependencies between different flood drivers to predict the likelihood of compound flooding events, providing insight into how extreme weather conditions can converge to heighten flood risks. Friederichs et al. (2018) illustrate that employing separate post-processing techniques on the tail distribution of heavy precipitation, specifically using Extreme Value Theory (EVT), is effective in generating reliable and accurate forecasts for extreme precipitation events, particularly during the summer when such events are frequently produced by mesoscale convective systems. In the application of RVT, Tabari (2021) observed that while both the Block-Maxima (BM) and Peak-Over-Threshold (POT) methods of EVT reflect a shift in intensity over time, they differ in the extent of this change, with the disparity increasing as the extremity of events rises. This research implies that although the WRF model might underestimate precipitation levels when EVT is applied, the correlation between flooding and rainfall remains consistent.

Johansson et al. (2003) developed a regression model to study the impact of wind and topography on precipitation distribution in Sweden and combined it with statistical analysis. The primary goal was to determine whether statistical relationships could describe the relationship between precipitation, airflow, and topography. However, they also indicate that the model could be enhanced by incorporating wind direction as a variable. Although machine learning has made significant improvements in weather forecasting, accurately predicting such events remains difficult due to the complexity of their interrelationships (Maina et al., 2023). As floods are influenced by multiple drivers that are often statistically interdependent, not counting these dependencies can result in the underestimation of flood risk and the inadequate design of flood control structures (Jane et al., 2020). Thorsten (2006) suggests that copulas can describe or model the interdependencies between random variables, making them a critical tool in modelling dependencies between different hazards in multi-hazard risk assessment.

Traditional multivariate probability distributions often require all marginal distributions to be of the same type (e.g., Gaussian), which may not always be appropriate for the data at hand, thus limiting the model's flexibility and accuracy. For example, Loganathan et al. (1987) assumed the marginal distributions between extreme tides and the freshwater flows they led were from Gaussian distribution and applied a bivariate Gaussian distribution to model these variables. They found that this traditional assumption of statistical independence between streamflow and tidal heights could lead to underestimation of flow and tidal values.

Copula theory, which allows for modelling dependencies between multiple variables, has been widely applied in risk modelling for both natural and financial risks and is increasingly used in multi-hazard scenarios. In the financial sector, Low et al. (2013) applied a family of copula models (Clayton canonical vine copula) to forecast the return on investment portfolios, finding that copulas work reasonably when dealing with portfolios with higher dimensions. In natural hazard modelling, a copula-based approach for flood risk analysis has been developed to assess the probabilistic risk of floods by analysing the relationship between volume, duration of the flood hydrograph, and peak flow (Ganguli et al., 2012).

Ming et al. (2022) describe using Copula functions, particularly the Clayton copula, to quantify the joint probability distributions of compound flooding events involving rainfall, river flow, and storm surges, offering a more nuanced understanding of how these hazards interact, leading to more accurate risk assessments. Similarly, Ming's study highlights the application of Monte Carlo simulations to generate stochastic multi-hazard events based on these copula-derived joint distributions, further supporting the probabilistic nature of multi-hazard modelling.

Shiau and Modarres (2009) use copulas to analyse the relationships between these factors within a drought model and conclude that frequency, duration, and severity are three relevant factors contributing to every drought event. On the issue correlated to extreme value dependence, t copula analysis is particularly useful in this situation (Demarta et al., 2007). In the exploration of advanced statistical methods for weather prediction, the latest approach has been presented using D-vine copula-based post-processing for wind speed forecasts, offering a refined technique for enhancing the accuracy of wind speed predictions. As a result, copula analysis can significantly model dependency structures between flood levels, wind speed, and rainfall to improve flood prediction (Jobst et al., 2023).

Building on this, numerical models like hydrodynamic models allow for simulating flood dynamics under varying multi-hazard conditions. These models, integrated with copula-based joint probability distributions, are essential for capturing the dynamic processes of hazards, particularly in urban areas where interactions between different hazards can significantly exacerbate impacts. Li et al. (2019) extend this approach by exploring the use of t-copulas and other types of copulas to model the dependency structures in water accessibility, demonstrating how different copula types can be applied depending on the environment's nature and the required dependency level.

Additionally, Bachner et al. (2023) emphasise the combined use of statistical models and simulations, advocating for an approach that integrates probabilistic assessments with dynamic simulations to address the complexities of cascading hazards, such as those observed in Austria, where socioeconomic factors compound flood risks. Finally,

Zscheischler et al. (2018) discuss the application of joint probability approaches, which are crucial for assessing the risk of concurrent hazards like storm surge and extreme rainfall, further supporting the necessity of combining copula theory with advanced simulation techniques for comprehensive multi-hazard risk assessments.

In the hazard model proposed in this paper, we will employ an EVT distribution to model wind speed, rainfall, and flood data for each variable. The parameters calculated from these distributions will then be used as marginal distributions in copula analysis to determine the most suitable copula types. After obtaining the best fitting copulas for our data, the model can be interpreted by inputting wind speed or rainfall levels to estimate the joint annual exceedance probability and corresponding flood levels.

2.3. Vulnerability Modelling

Vulnerability in the context of flood risk is a multidimensional concept that reflects the potential for loss due to the interplay of physical and social factors. According to Cutter (1996) and Cutter et al. (2003), vulnerability is determined by natural exposure, referred to as physical and socio-demographic characteristics or social vulnerability. Physical vulnerability involves the geographical location and environmental conditions that make certain areas susceptible to flood hazards, indicating where a hazard is likely to occur and which assets are exposed (Cutter et al., 2003; Hufschmidt, 2011). On the other hand, social vulnerability arises from societal factors that influence a community's capacity to prepare for, respond to, and recover from flood events. These factors include unequal access to resources, varying levels of power and influence, settlement patterns, and the broader social order (Morrow, 1999). For example, residents in flood-prone areas may suffer significant economic losses; however, those with higher income levels or adequate insurance coverage are better equipped to recover from such events.

Turner et al. (2003) further elaborate on this dual nature of vulnerability by integrating these physical and social dimensions into a broader framework for vulnerability assessment. This framework highlights the interaction between human and environmental systems,

emphasising that vulnerability is shaped not only by exposure to hazards but also by the underlying social structures and processes that influence how communities experience and respond to such risks. Several statistical methodologies have been employed in vulnerability assessments to manage the complexity of the data and ensure that the indices or models developed are robust and reliable. Poudyal et al. (2012) and subsequent studies by Mentzafou et al. (2016) have applied various techniques to develop these models. Common methods include normalizing indicators and analysing inter-indicator correlations to identify relationships between different vulnerability indicators.

Jongman et al. (2015) developed a global model to assess flood vulnerability by integrating socioeconomic factors and infrastructure data. Their model highlighted the varying levels of vulnerability across different regions, underscoring the need for localised risk assessments. Thieken et al. (2016) proposed a vulnerability model that accounts for building characteristics, flood depth, and duration, enabling detailed loss estimation at the building level and providing valuable information for risk management and insurance purposes. Merz et al. (2021) focused on the use of multi-variable models in flood vulnerability assessments, comparing these models to traditional single-variable approaches. They demonstrated that multi-variable models, which include parameters such as water velocity, building type, and occupancy, significantly outperform single-variable models. This study reinforced the importance of comprehensive data integration for accurate flood damage prediction.

Traditional vulnerability models often do not account for financial losses, although Bakkensen et al. (2016) and Sharifi (2016) summarise and normalise socioeconomic indicators such as language, disability, minority status, and other factors that influence a community's vulnerability to natural hazards, resulting in a single score to guide policymakers and flood planners. These researches proposes a flood vulnerability model that primarily focuses on financial losses to provide flood emergency planners with sufficient resources for rebuilding the community. Current approaches to estimating flood vulnerability recognise the critical role of financial losses as a key vulnerability indicator. Financial losses directly impact a community's ability to recover from flood events. Floods

can disrupt essential facilities and services, which are crucial for maintaining a community's quality of life and economic well-being (Scheuer et al., 2010).

In general insurance pricing, machine learning algorithms such as Generalised Linear Models (GLM) are used in severity modelling (Goldburd et al., 2020; Pietro Parodi, 2015). In the application of machine learning algorithms, they involve selecting model complexity through cross-validation and controlling it with one or more hyperparameters. Unlike linear and parametric methods, it is non-linear and non-parametric, offering flexibility that often results in strong predictive performance, though these models are generally less interpretable (Molnar et al., 2020). Therefore, the proposed model uses a GLM to address the gap in incorporating the financial component into vulnerability modelling. Additionally, the random components (probability distribution of the target variable) are assumed to belong to a member of the exponential family of distributions, such as the gamma distribution (Anderson et al., 2007).

Beyond the financial sector, machine learning techniques have been applied to flood risk analysis. A machine learning based flood hazard assessment and risk mapping framework is presented via optical remote sensing and GIS data. The framework employs explainable machine learning techniques, particularly Random Forest, to assess flood hazard levels dynamically. By integrating satellite imagery with machine learning, this approach provides a comprehensive method for flood mapping, establishing a precedent for similar applications in flood vulnerability analysis (Antzoulatos et al., 2022). However, there is a need for more comprehensive models that integrate all components of risks and vulnerability factors, as existing models often lack detailed vulnerability assessments (Aerts et al., 2018).

It is undeniable that advanced machine learning or deep learning algorithms statistically outperform GLM in predictive accuracy. However, Molnar et al. (2020) caution that several pitfalls can arise in the interpretation of advanced machine learning models, particularly when they are applied inappropriately. These include misapplying the model to the wrong context, overlooking dependencies between features, disregarding uncertainty in estimates, and misunderstanding causality when using illustrative graphs. Such issues can

lead to significant errors, especially when users mistakenly assume that a single machine learning model can provide consistent interpretations across all contexts. The challenge lies in selecting an interpretation method that offers accurate predictions while effectively translating detailed model results into comprehensible explanations. Birkland (2006) also discusses how policymakers often operate with bounded rationality and limited information, making it difficult for them to fully understand and utilise complex catastrophe models. Therefore, using GLM for vulnerability modelling in this paper strikes a balance between complexity and interpretability.

Notably, companies such as Verisk have developed commercially used wind-related flood hazard models to provide their clients, particularly those in the insurance and reinsurance industries, with hurricane risk solutions (Grenier et al., 2020). However, these companies charge high annual premiums for their services, and finding an open-access or academically used catastrophe model that combines hazard and vulnerability models with financial losses is rare. This paper aims to construct a model that integrates modules from sections 2.2 and 2.3 to develop an extreme wind-related flood hazard framework, allowing users to predict financial losses by inputting rainfall level or windspeed data.

3. Methodology

3.1. Research Design

The primary objective of this research is to develop a catastrophe model framework that integrates a hazard model and a vulnerability model to estimate the financial impact of extreme weather events, specifically focusing on flood-related damages in the US Golf Coastal. The research follows a quantitative, model-based approach, utilising statistical methods to capture the relationships between extreme environmental conditions and the resulting damage to building infrastructure.

This model aims to address the need for reliable predictive tools in disaster risk management by integrating two key components: a hazard model and a vulnerability model.

Hazard Model Objective: The hazard model is designed to simulate the intensity and joint occurrence of extreme weather events, such as rainfall, wind speed, and flood depth. By applying advanced statistical methods, particularly Extreme Value Theory (EVT) and copula analysis, the model generates a range of potential future hazard scenarios for the region. These scenarios are crucial for understanding the potential severity and frequency of extreme events in the target area. EVT and copula analysis are selected for this research because EVT provides a robust framework for modelling the tail behaviour of probability distributions, which is essential for analysing rare and extreme events such as severe floods and hurricanes. Copula functions are employed to model the dependency structure between multiple variables, such as rainfall, wind speed, and flood depth. Traditional statistical methods often struggle to accurately capture the distribution of extreme values, leading to an underestimation of risk. EVT addresses this issue by focusing on the tail behaviour of probability distributions, allowing for a more precise estimation of the intensity and likelihood of extreme weather events. Copula analysis complements EVT by modelling the joint distribution of multiple variables, even when they exhibit non-linear or tail dependencies, providing a nuanced and realistic assessment of

multivariate risks. This approach is essential in multivariate extreme value analysis, where understanding the simultaneous occurrence of multiple hazards enhances the accuracy of risk assessment. Copulas enable the model to capture complex dependencies that traditional correlation measures might overlook, thereby offering a more detailed and realistic representation of the risk.

Vulnerability Model Objective: The vulnerability model aims to quantify the relationship between the simulated hazard scenarios and the resulting building damages. This component of the research utilises a Generalised Linear Model (GLM) approach, incorporating various building and environmental characteristics that influence the extent of damage. The goal is to develop a model that can accurately predict the financial losses associated with specific hazard scenarios. The GLM framework is well-suited for modelling the relationship between hazard scenarios and building damages due to its ability to accommodate different types of response variables (e.g., continuous, categorical, binary) and error distributions (e.g., Gamma, log-normal). This flexibility is crucial given the diverse nature of the data involved, which includes financial losses, building characteristics, and environmental factors. Additionally, GLM allows for including both linear and non-linear relationships, making it ideal for capturing the complex interactions between the predictors and the response variable, in this case, the extent of building damage. The forward selection method employed in this research ensures that only the most significant predictors are included in the final model, thereby enhancing its predictive accuracy.

The integration of the hazard and vulnerability models is central to the research design, enabling the simulation of potential losses under various extreme weather scenarios. The combined model is then evaluated using cross-validation techniques to ensure its reliability and predictive accuracy. The research design is iterative, with feedback from model evaluations used to refine both the hazard and vulnerability components.

To achieve these objectives, the research will proceed through the following key steps:

- **Data Collection and Preprocessing (3.2):** Collect and preprocess data from the National Flood Insurance Program (NFIP) and meteorological records, ensuring the datasets are suitable for the development of both hazard and vulnerability models.

- **Hazard Model Development (3.4):** Implement the hazard model using EVT and copula analysis to simulate extreme weather scenarios that could affect the US Gulf Coastal region.
- **Vulnerability Model Development and Validation (3.5):** Develop the vulnerability model using GLM to estimate the financial impact of the simulated scenarios on building infrastructure. Apply cross-validation techniques to assess the reliability and predictive accuracy of the integrated model, refining the model as necessary based on the validation results.
- **Model Integration (3.6):** Integrate the hazard and vulnerability models to form a unified catastrophe model capable of predicting potential losses under extreme weather conditions and various building characteristics.

3.2. Data Collection and Description

3.2.1. Data Sources

The data source for this research is the Federal Emergency Management Agency (FEMA) National Flood Insurance Program (NFIP) claims dataset and National Centers for Environmental Information (NCEI) data. NFIP dataset from FEMA contains detailed records of insurance claims made by policyholders affected by flood events across the United States, including Texas, Louisiana and Mississippi. The dataset contains a variety of variables, including the cause of damage, state, date of loss, building damage amount, flood event names, and other related information. These records will be used in developing the vulnerability model.

In addition to the NFIP claims data, meteorological data was collected from the National Centers for Environmental Information (NCEI). The weather station data, provides daily measurements of rainfall level (precipitation) and windspeed. These data points will be used to build the hazard model.

After filtering the database, the available data are as follows:

- Hazard Model: Texas: 9,170 observations from 1998-2024;
- Louisiana: 9,956 observations from 1996-2023;
- Mississippi: 9,137 observations from 1998-2023.

Vulnerability Model:

- Texas: 13,261 observations from 1998-2019;
- Louisiana: 4,430 observations from 1997-2021;
- Mississippi: 254 observations from 1998-2016.

Table A1 shows the variables extracted from the flood insurance claims data and meteorological data and their descriptions in detail. These variables will construct the dataset we use in the model-building process.

3.3. Data Preprocessing and Visualisation

Once we obtain and collect data from flood insurance claims and meteorological data, we calculate the water depth from the insurance claim database by getting the averages for a single day.

$$Average\ Water\ Depth\ (flood\ level) = \frac{sum\ of\ water\ depth\ in\ one\ day}{number\ of\ claims\ on\ that\ day}$$

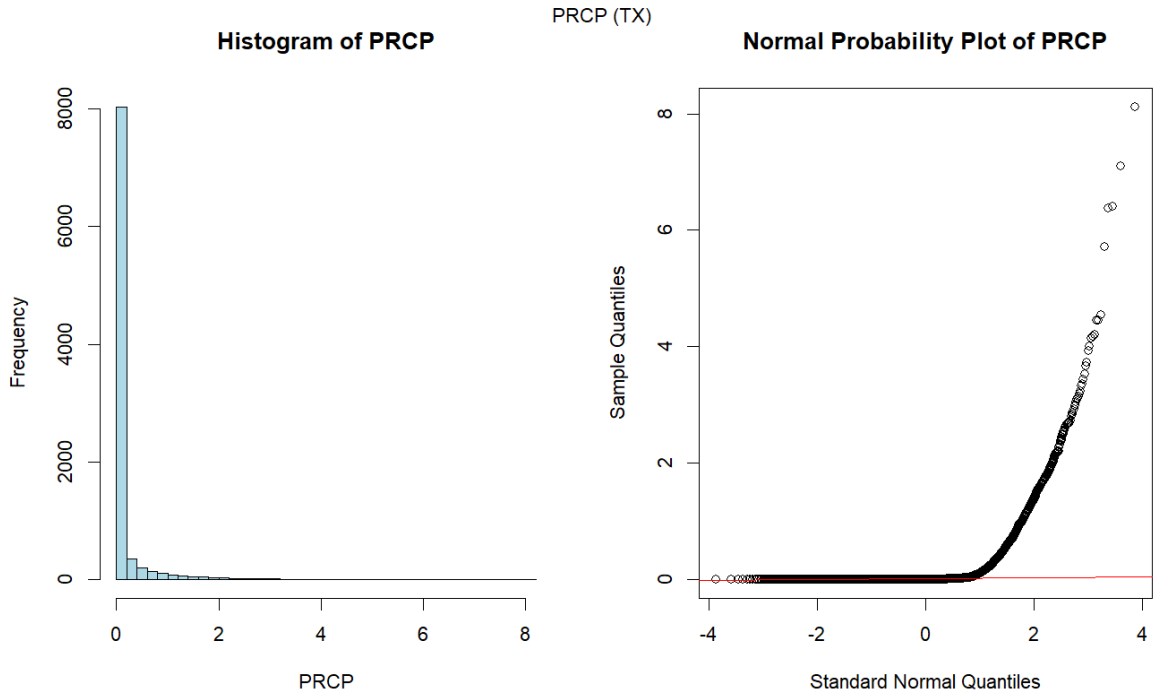

Figure 3.1: Distribution and Normality Assessment of Rainfall Level (Texas)

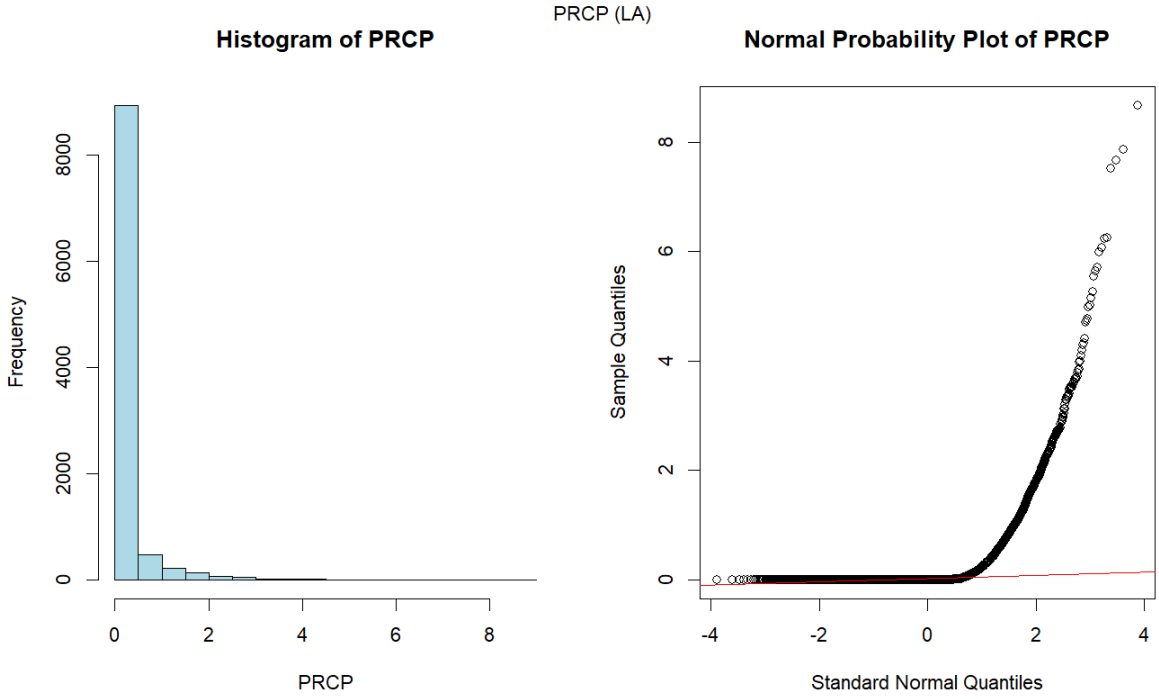

Figure 3.2: Distribution and Normality Assessment of Rainfall Level (Louisiana)

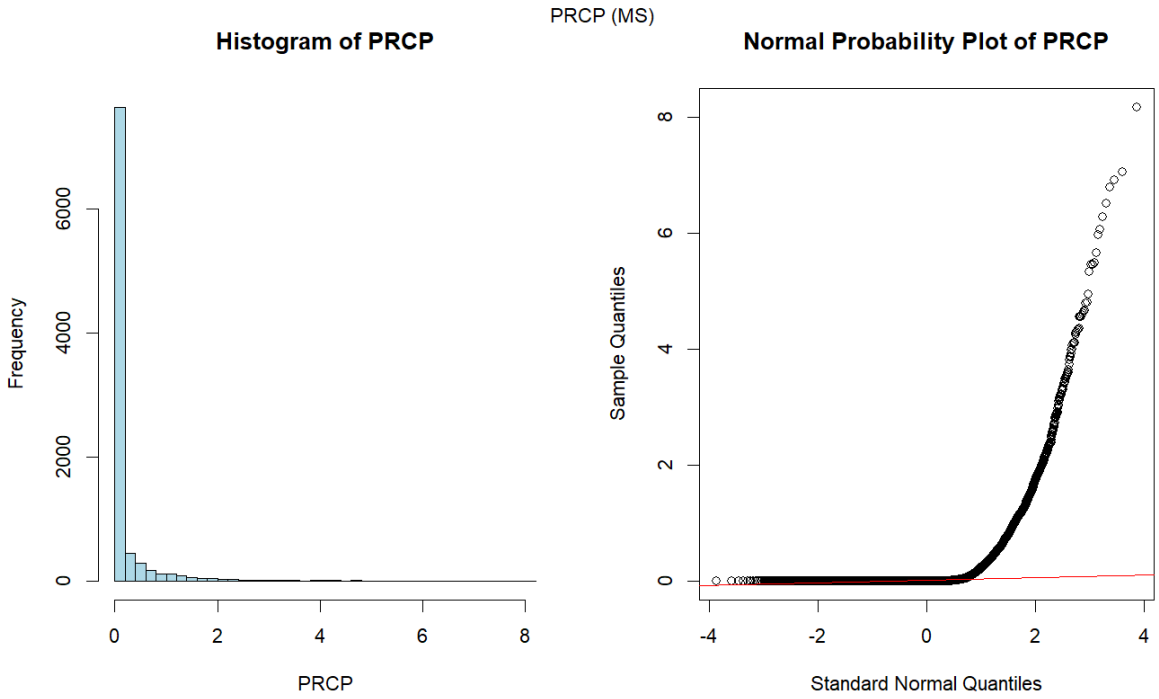

Figure 3.3: Distribution and Normality Assessment of Rainfall Level (Mississippi)

Figures 3.1, Figure 3.2, and Figure 3.3 show the distribution and normality of rainfall levels in Texas, Louisiana and Mississippi. Across all three states, the rainfall data is highly right-skewed, with a significant number of observations clustered near zero and a long tail extending towards higher values. This indicates that most recorded rainfall events are low, but there are occasional extreme events. The Q-Q plots confirm this right skewness, showing a significant departure from normality, particularly in the upper quantiles where the data deviates sharply from the expected normal distribution line.

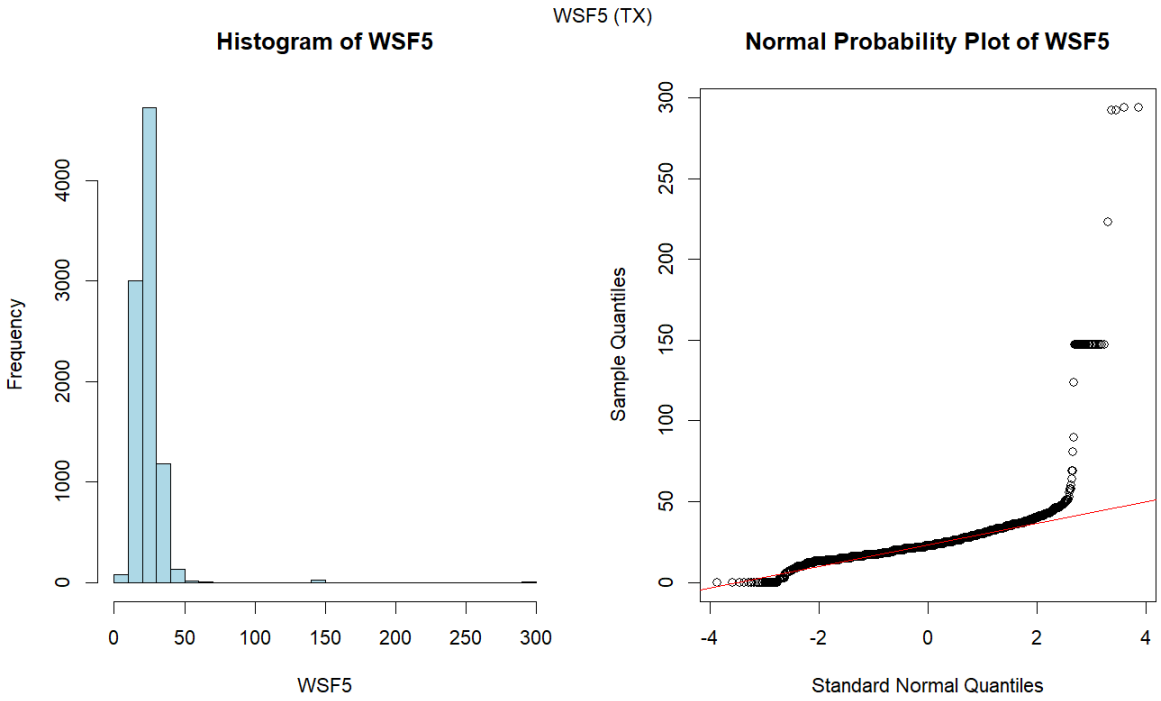

Figure 3.4: Distribution and Normality Assessment of Windspeed (Texas)

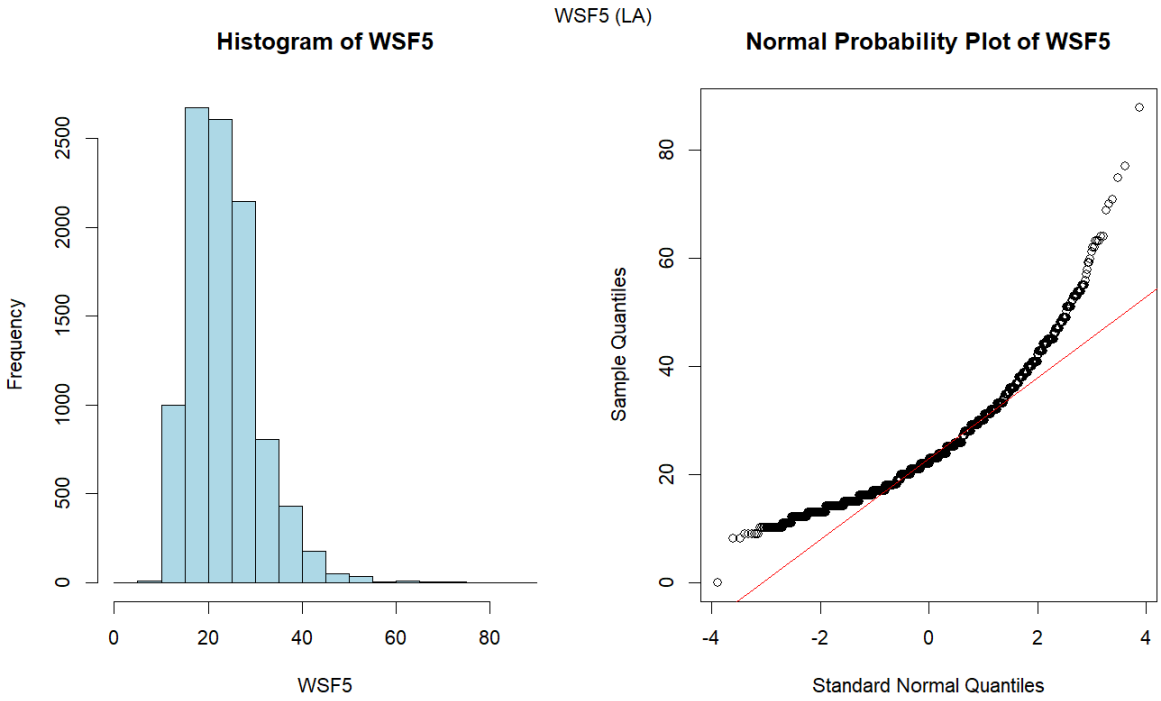

Figure 3.5: Distribution and Normality Assessment of Windspeed (Louisiana)

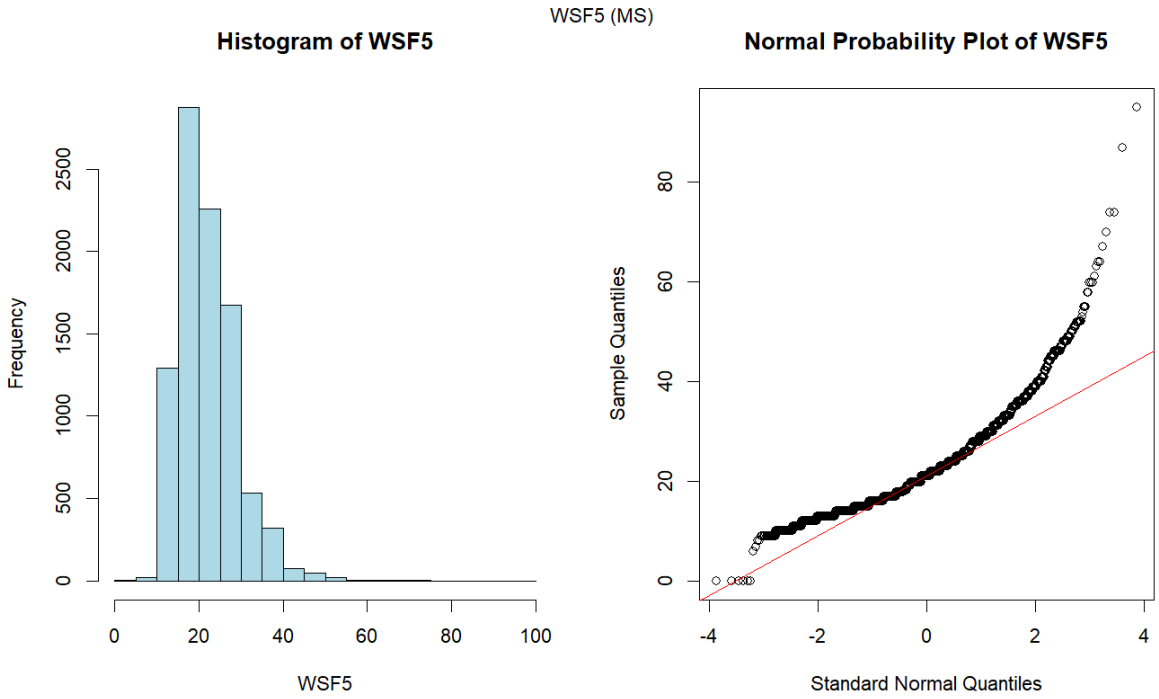

Figure 3.6: Distribution and Normality Assessment of Windspeed (Mississippi)

Windspeed data across Texas, Louisiana, and Mississippi (Figures 3.4, 3.5, and 3.6) is also right-skewed, although the skewness is less extreme than that observed in the rainfall data. The histograms show that most windspeed observations are lower, with fewer instances of high wind speeds. The Q-Q plots for windspeed suggest a moderate deviation from normality, particularly in the upper tail, which is characteristic of right-skewed data. However, this deviation is less obverse than in the rainfall data.

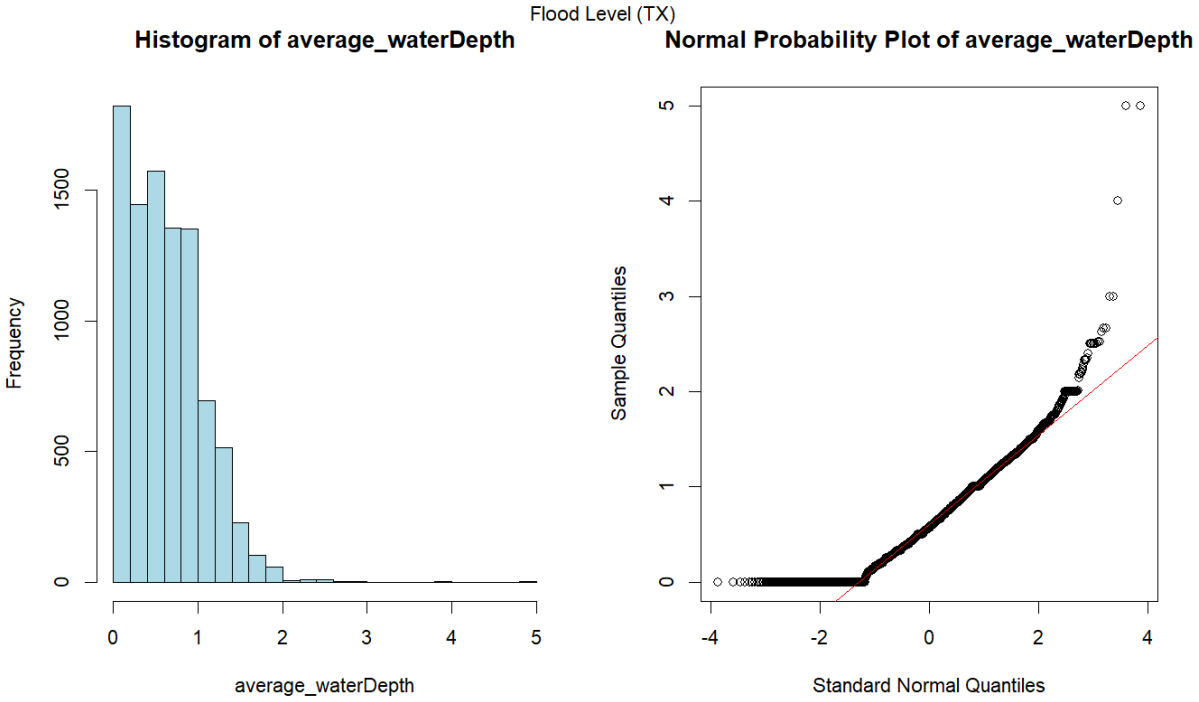

Figure 3.7: Distribution and Normality Assessment of Flood Level (Texas)

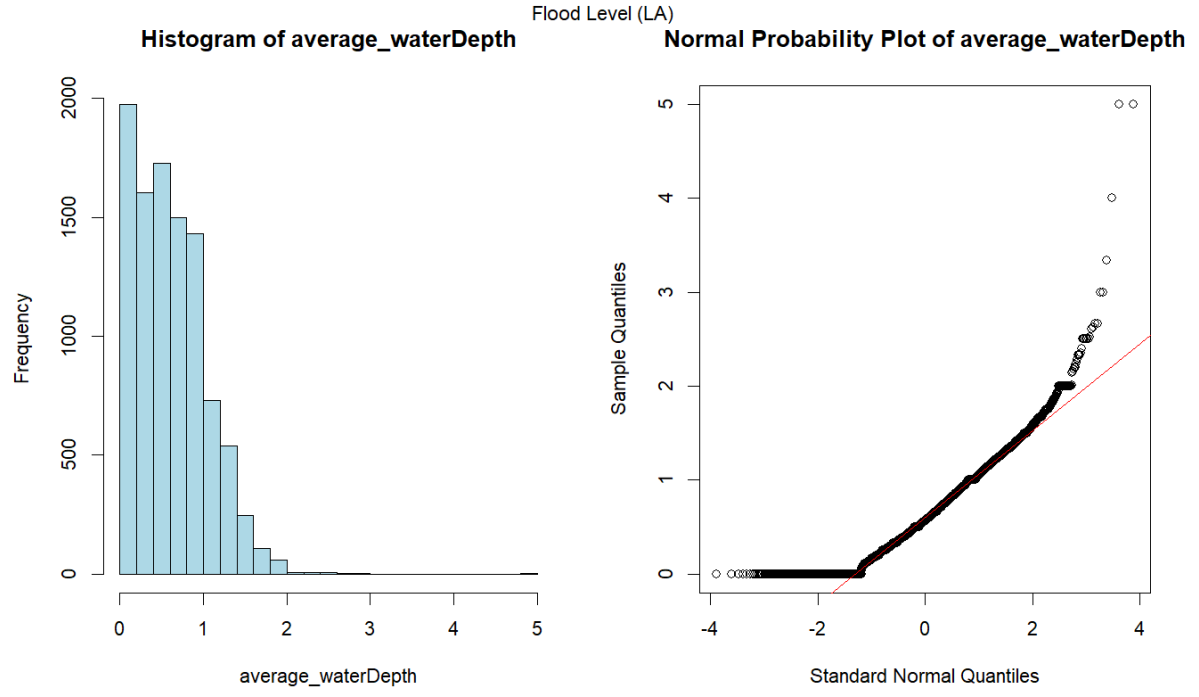

Figure 3.8: Distribution and Normality Assessment of Flood Level (Louisiana)

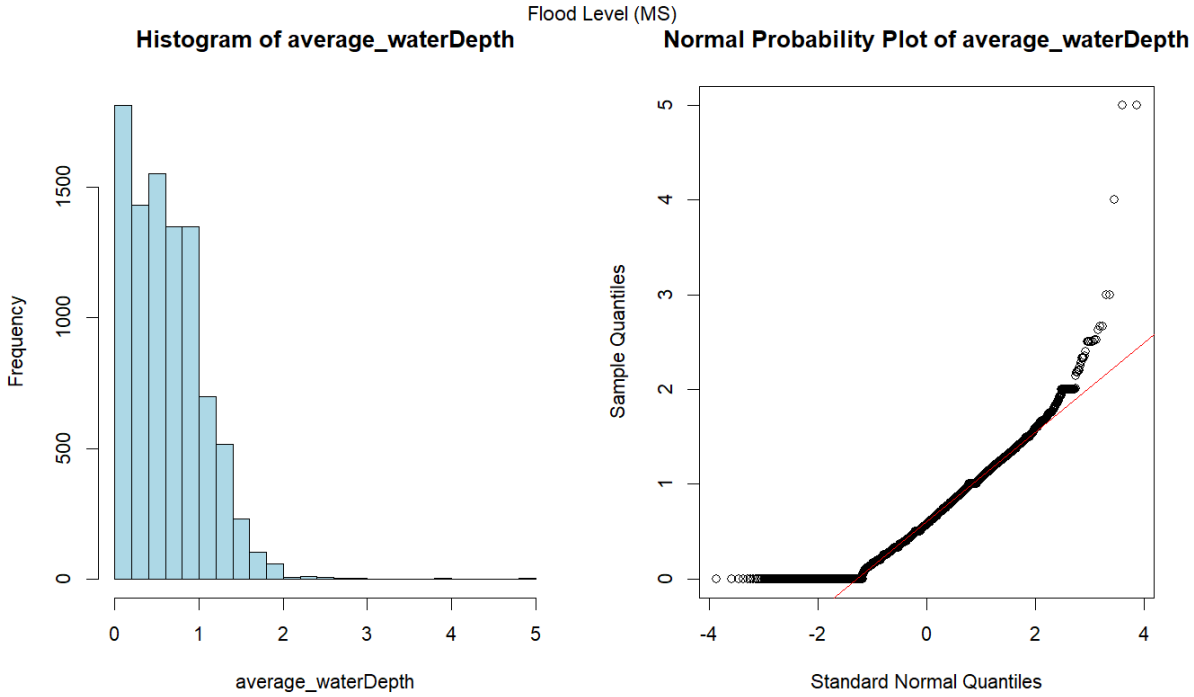

Figure 3.9: Distribution and Normality Assessment of Flood Level (Mississippi)

Similar to rainfall, the flood level data (Figures 3.7, 3.8, and 3.9) show a right-skewed distribution, with most observations at lower flood depths and fewer instances of higher flood levels. The Q-Q plots again demonstrate a significant deviation from normality, particularly in the upper quantiles, indicating that the data contains some extreme flood events.

All nine figures consistently demonstrate that the rainfall, windspeed, and flood level data are not normally distributed and exhibit right skewness. This non-normality suggests that traditional statistical methods that assume normality may not be appropriate without data transformation or the use of alternative modelling techniques. Given the skewness and the presence of extreme values, these data are well-suited for modelling using the Generalised Extreme Value (GEV) distribution and the Generalised Pareto Distribution (GPD), which are designed to handle extreme events and heavy-tailed distributions.

In addition, the consistency of distribution patterns across Texas, Louisiana, and Mississippi indicates that these variables (rainfall, windspeed, flood levels) exhibit similar characteristics across different geographic regions. This consistency likely reflects underlying environmental factors common to these states, such as climate patterns or geographic features associated with the US Gulf Coastal Region.

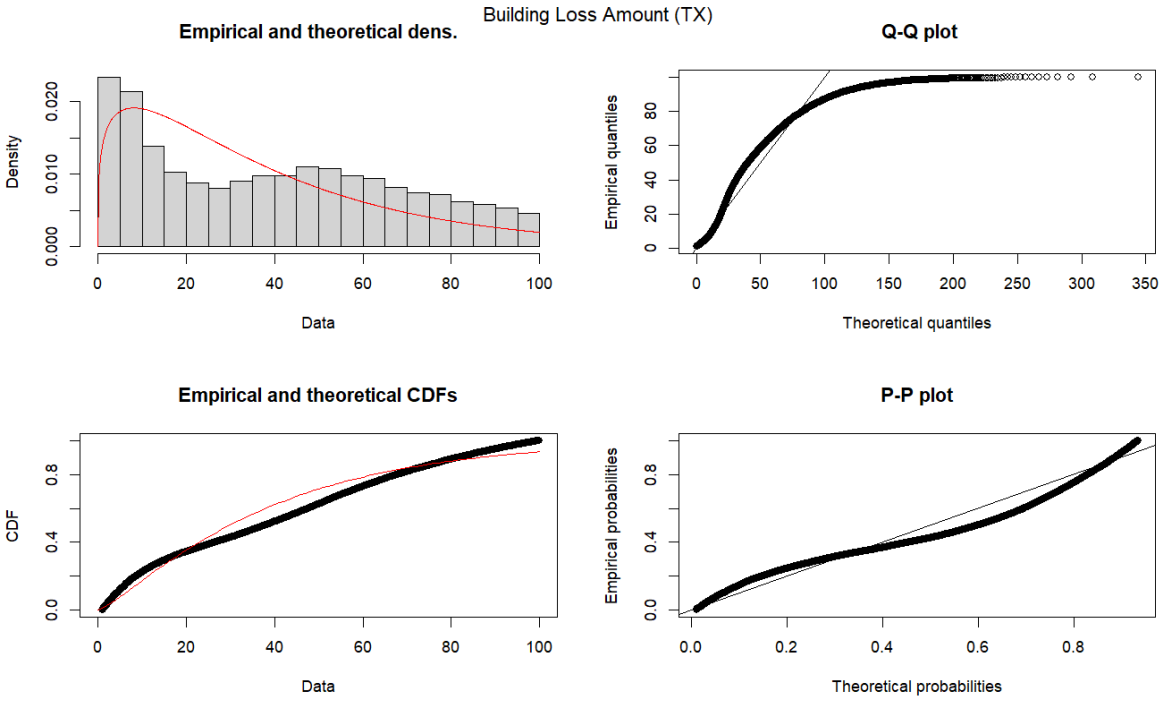

Figure 3.10: Gamma Fit Diagnostics for Building Loss Amount (Texas)

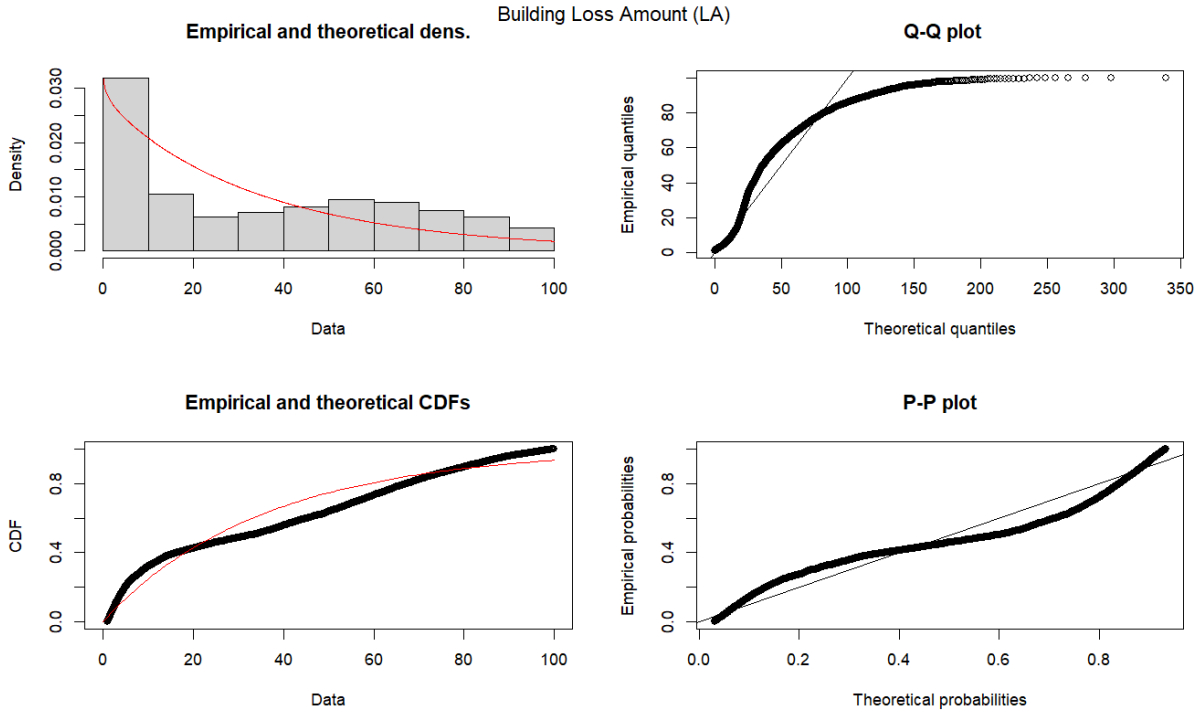

Figure 3.11: Gamma Fit Diagnostics for Building Loss Amount (Louisiana)

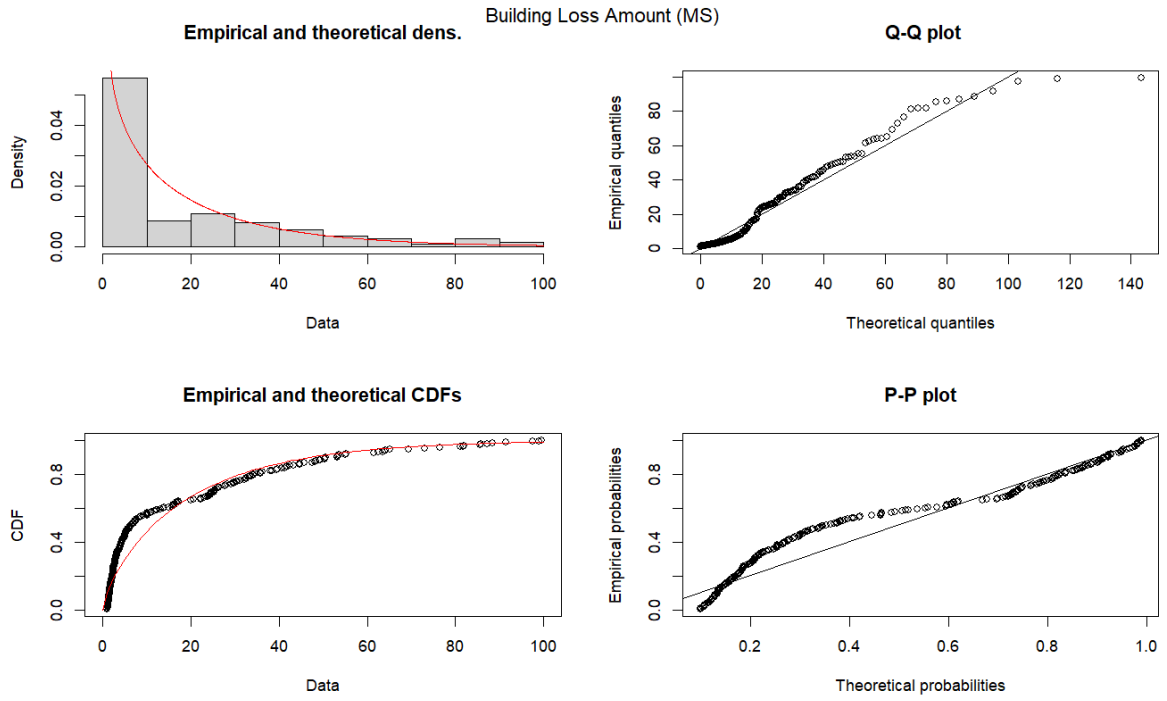

Figure 3.11: Gamma Fit Diagnostics for Building Loss Amount (Mississippi)

Figures 3.9, Figure 3.10, and Figure 3.11 are the Gamma fit diagnostic plots for the building loss amounts in Texas, Louisiana, and Mississippi. These diagnostic plots suggest that while the Gamma distribution provides a reasonable fit for the building loss amount data across all three states, there might be potential limitations in capturing the extreme values, especially in Texas. The tails in Q-Q and P-P plots show this potential underestimation, where deviations from the theoretical line occur at higher quantiles and probabilities. Although these differences suggest that extreme losses might not follow a Gamma distribution, overall, the Gamma distribution is an acceptable fit for flood loss data.

3.4. Hazard Model Development

3.4.1. Extreme Value Theory

3.4.1.1 Block Maxima Approach

The block maxima approach is designed to model extreme values, and it can be used to model extreme values (events) using environmental data. This approach divides the complete dataset into non-overlapping blocks of a fixed size by a unit of time length (such as years) and extracts the maximum value in each block. These extracting values (maximum values in each block) are then used to estimate the parameters in extreme value distribution. This approach is widely used in estimating the parameters for Generalised Extreme Value (GEV) distribution. Also, in this study, we use the Maximum Likelihood (ML) method to estimate the distribution parameters since, compared to the probability Weighted Moments (PWM) method, the ML method is more flexible in parameterisation. (Coles & Casson, 1998; Davison & Smith, 1990)

3.4.1.2 Generalised Extreme Value Distribution (GEV)

The GEV distribution is a family of continuous probability distributions used to model the maximum (or minimum) of a set of independent, identically distributed random variables.

The cumulative distribution function (CDF) and the probability density function (PDF) of the GEV distribution are expressed as follows:

$$F(y) = \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \quad \text{for } 1+\xi\left(\frac{y-\mu}{\sigma}\right) > 0; \quad \textit{Equation 3.1}$$

$$f(y) = \frac{1}{\sigma}\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \exp\left(-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right), \quad \text{for } 1+\xi\left(\frac{y-\mu}{\sigma}\right) > 0; \textit{ Equation 3.2}$$

Where:

μ is the location parameter, which represents the location of the distribution.

σ is the scale parameter, which determines the spread of the distribution.

ξ is the shape parameter, which determines the shape of the distribution. When $\xi = 0$, the GEV distribution becomes the Gumbel distribution; when $\xi > 0$, it represents the maximum distribution; when $\xi < 0$, it represents the minimum distribution.

3.4.1.3 Peak over Threshold Approach

The peak over threshold approach is like the block maxima approach. However, instead of getting the maximum value in each block, the peak over threshold approach sets a threshold value and exacts all values above this threshold. The data is modelled using the Generalised Pareto Distribution (GPD) beyond the threshold. The GPD is used to estimate the tail behaviour of the dataset with heavy tails, and the parameters of the distribution are estimated using the maximum likelihood method.

3.4.1.4 Generalised Pareto Distribution

The GPD is another distribution commonly used in extreme value analysis, particularly in the threshold approach. The GPD is a continuous probability distribution that can be used to model the excesses over a threshold of a random variable.

The cumulative distribution function (CDF) and the probability density function (PDF) of the GPD distribution are represented by:

$$H(y) = 1 - \left(1 + \xi\frac{y}{\sigma}\right)^{-\frac{1}{\xi}}, \quad \text{for } y > 0 \text{ and } \left(1 + \xi\frac{y}{\sigma}\right) > 0; \quad \textit{Equation 3.3}$$

$$h(y) = \frac{1}{\sigma}\left(1 + \xi\frac{y}{\sigma}\right)^{-\frac{1}{\xi}-1}, \quad \text{for } y > 0 \text{ and } \left(1 + \xi\frac{y}{\sigma}\right) > 0; \quad \textit{Equation 3.4}$$

Where:

μ is the location parameter, which represents the threshold.

σ is the scale parameter, which determines the spread of the distribution.

ξ is the shape parameter, which determines the shape of the distribution. When $\xi > 0$, it represents a heavy-tailed distribution, and when $\xi = 0$, it represents an exponential distribution.

3.4.2. Copula Analysis

In the context of multivariate hazard modelling, it is important to understand the dependency structure between different variables when understanding the relationship of the hazards, which may occur simultaneously or be influenced by common underlying factors. To understand this relationship, copula functions provide a framework for modelling the dependency structure between variables, allowing for an accurate representation of dependencies. In this study, it provides the correlations between the extreme values of the variables.

The copula function is a type of multivariate probability distribution function that links the marginal distributions of variables to a multivariate distribution, with each marginal being uniform over the interval [0,1] (Nelsen, 2006). This approach is particularly useful in statistical modelling as it allows for the independent capture of the dependency structure between variables, regardless of their marginal distributions.

Any multivariate joint distribution can be represented using its marginal distributions and a copula that captures the dependency structure between the variables by Sklar's Theorem. Specifically, for a pair of random variables *X* and *Y* with a joint distribution function *H(x,y)* and marginal distributions *F(x)* and *H(y)*, there exists a copula *C(u,v)* such that: H(x,y)=C(F(x),H(y)). In extending this concept to higher dimensions, consider a set of random variables $X_1, X_2, \ldots, X_d$ with corresponding marginal distribution functions $F_i(x_i)$, $i=1,\ldots,d$. The joint distribution function for these variables is denoted by $J(x_1, \ldots, x_d)$. According to Sklar's theorem (Sklar, 1959), a copula C: $[0,1]^d \rightarrow [0,1]$ exists such that:

$$F(x_1, \ldots, x_d) = C\big(F_1(x_1), \ldots, F_d(x_d)\big)$$

When the marginal distributions $F_i(x_i)$ are all continuous, the d-dimensional copula C is unique and can be expressed as:

$$C(u_1, \ldots, u_d) = F\Big(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\Big)$$

where $F_i^{-1}(u_i)$ denotes the inverse of the marginal distribution function $F_i(x_i)$

Copula functions used in this paper are following (Nelsen, 2011; Bezak et al., 2014):

t Copula Function:

$$C_{\nu,P}(u_1, u_2, \ldots, u_d) = t_{\nu,\mathrm{P}}\big(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_d)\big) \quad \textit{Equation 3.5}$$

Gumbel Copula Function:

$$C_\theta(u_1, u_2, \ldots, u_d) = \exp\left[-\left(\sum_{i=1}^{d}(-\log u_i)^\theta\right)^{\frac{1}{\theta}}\right]; \quad \textit{Equation 3.6}$$

Frank Copula Function:

$$C_\theta(u_1, u_2, \ldots, u_d) = -\frac{1}{\theta}\log\left[1 + \frac{\prod_{i=1}^{d}(e^{-\theta u_i}-1)}{(e^{-\theta}-1)^{d-1}}\right]; \quad \textit{Equation 3.7}$$

Clayton Copula Function:

$$C_\theta(u_1, u_2, \ldots, u_d) = \left[\sum_{i=1}^{d} u_i^{-\theta} - (d-1)\right]^{-\frac{1}{\theta}}; \quad \textit{Equation 3.8}$$

The probabilistic density function of the copula C, often referred to as the copula density, can be denoted as c(u). This copula density function c(u) plays a crucial role in capturing the strength and nature of dependencies between the random variables, particularly in fields where understanding joint behaviour can be calculated by the partial derivative of the copula function.

Copula density function:

$$c(u_1, u_2, \dots, u_d) = \frac{\partial^d C(u_1,u_2,\dots,u_d)}{\partial u_1 \partial u_2 \dots \partial u_d}; \quad \textit{Equation 3.9}$$

In this paper, we are using three copulas from the Archimedean family (Gumbel, Frank and Clayton) and one from the elliptical copula family (t copula). We used the Maximum Pseudo-Likelihood (MPL) method to estimate the copula parameters. It fits the copula model by maximising the log pseudo-likelihood function (Genest et al., 1995), which is defined as follows:

Log Pseudo-likelihood function:

$$log\ L(\theta;\ u_1, u_2, \dots, u_n) = \sum_{\{i=1\}}^{n} \log c_\theta\ (\boldsymbol{u_i}) \quad \textit{Equation 3.10}$$

For each copula function, we first apply *Equation 3.9* to derive the copula density function. After we get the copula density function, we use *Equation 3.10* to maximise the log pseudo-likelihood function, allowing us to estimate the copula parameters.

3.5. Vulnerability Model Development

3.5.1. Machine Learning Approach (Generalised Linear Model)

Nelder and Wedderburn (1972) introduced the Generalised Linear Model (GLM), a significant advancement in statistical modelling that extends the traditional linear regression framework to accommodate various types of response variables and error distributions. The GLM framework enhances flexibility by allowing the model to handle different data types, making it suitable for a broad range of applications.

A Generalised Linear Model (GLM) is composed of three primary components:

- Random Component: This denotes the probability distribution of the response variable, which is part of the exponential family of distributions (e.g., Poisson, Gamma).
- Systematic Component: This component is the linear predictor, consisting of a linear combination of the explanatory variables x and their corresponding regression coefficients β.
- Link Function: The link function serves to connect the random component with the systematic component, establishing a relationship between the expected value of the response variable (random component) and the linear predictor (systematic component).

We are using forward search to identify the most effective covariates from all available options, aiming to construct the most efficient GLM model for estimating losses associated with a single building.

In ***forward search:***

p represents the significance level that determines when to stop the forward selection process. In here, $p = 0.05$

Step 1:	Fit the model with no covariates (intercept only)
Step 2:	Evaluate each potential predictor by adding it to the current model one at a time and fit the model.
Step 3:	Add the selected predictor to the model and remove it from the list of potential predictors.
Step 4:	Repeat Steps 2 to Step 4, adding one predictor at a time, until no remaining predictors have a p-value below p
End	

Once the most effective covariates have been identified, we construct GLM models using these selected variables and evaluate whether including interaction terms can enhance the model's accuracy. Following this, predictions are made based on the simulated scenario data.

3.6. Integration of Hazard and Vulnerability Models

Expected Loss for a single building is calculated by the formula:

$$Expected\ Annual\ Loss: Frequency \times Vulnerability \quad \textit{Equation 3.11}$$

Expected Annual Loss is defined as the expected financial losses for a single type of building at certain windspeed, rainfall, and flood thresholds.

Frequency is the exceedance probability of events exceeding certain thresholds of windspeed, rainfall, and flood levels.

Vulnerability in this paper is defined as the expected financial losses for a single type of building when events exceed certain thresholds of windspeed, rainfall, and flood levels.

3.7. Model Evaluation Techniques and Cross-Validation

3.7.1 AIC and BIC

The Akaike Information Criterion (AIC), introduced by Akaike in 1974, is a measure to evaluate the relative amount of information lost when a model is used to approximate the actual observed data. The AIC is calculated using the following formula:

$$AIC = -2\log maximum\ likelihood\ of\ the\ model + 2k$$

Where k is the number of estimated parameters in the model.

The Bayesian Information Criterion (BIC) is a model selection criterion based on Bayesian probability, which assumes a uniform prior distribution over the model space to estimate

the posterior probability of a model. Wit, Van Den Heuvel, and Romeijn (2012) formally define BIC as follows:

$$BIC = -2\log maximum\ likelihood\ of\ the\ model + k + ln(n)$$

Where k represents the number of estimated parameters in the model, and n donates the number of data points in observed data.

3.7.2 *k*-Fold Cross Validation

The *k*-Fold cross validation technique divides the dataset into *k* folds (subsets) of the same size *r* (Picard & Cook, 1984). In this study, we use 10 fold cross validation which indicates that k=10.

In ***k-Fold Cross Validation:***

Step 1:	dividing the data into k folds where each fold contains the same number of sample sizes *r*
Step 2:	fitting the model on remaining k-1 folds *(total number of observations – r)*
Step 3:	use the fitted model to predict the response variable on the excluded fold observations
Step 4:	calculate test Root Mean Square Error (RMSE)
Step 5:	repeat by excluding a different fold
End	

After all folds are chosen randomly and have been used as the test set once, the RMSE values are approximated by the averaged across all folds to provide an overall performance measure.

4. Results

4.1. Hazard Model Results

4.1.1. Block Maxima

4.1.1.1 GEV Parameter Estimation

Rainfall:

State	TX		LA		MS	
Parameters	Estimates	SE	Estimates	SE	Estimates	SE
Location	2.7224	0.2004	3.8615	0.2804	4.350	0.3447
Scale	0.9068	0.1658	1.1220	0.2386	1.579	0.2506
Shape	0.2686	0.1758	0.2635	0.2843	-0.311	0.1398

Table 4.1: Rainfall GEV Parameters and Standard Errors

From *Table 4.1*, Mississippi has the highest location parameter, suggesting that extreme rainfall events in Mississippi generally occur at higher rainfall levels than in Texas and Louisiana. Mississippi also exhibits the highest scale parameter, which suggests that extreme rainfall in Mississippi is more intense and variable than in Texas and Louisiana. The positive shape parameters for Texas and Louisiana suggest a heavy-tailed distribution, indicating that these states may experience more frequent extreme rainfall events with very high values. Conversely, Mississippi's negative shape parameter suggests an upper bound to the extreme rainfall values, implying less likelihood of exceptionally high extremes in this state compared to Texas and Louisiana.

The results indicate that Mississippi experiences higher and more variable extreme rainfall events but with an upper limit to the extremity of these events (as suggested by the negative shape parameter). In contrast, Texas and Louisiana tend towards more frequent extreme rainfall events with no apparent upper bound, highlighting potential concerns for more severe weather events in these states.

Wind Speed:

	TX		LA		MS	
	Estimates	SE	Estimates	SE	Estimates	SE
Location	54.932	3.3944	54.17716	1.6502	51.0761	1.565
Scale	15.721	4.7281	7.68311	1.2154	6.6287	1.427
Shape	1.054	0.2486	0.02496	0.1509	0.4208	0.232

Table 4.2: Windspeed GEV Parameters and Standard Errors

From *Table 4.2*, the Location parameter is highest in Texas, suggesting that the threshold for extreme wind speeds is slightly higher in Texas compared to Louisiana and Mississippi. This implies that Texas generally experiences higher wind speeds before they are classified as extreme. Texas's significantly higher scale parameter indicates much greater variability in extreme wind speeds. This suggests that Texas experiences a wider range of extreme wind speeds than Louisiana and Mississippi, where the wind speed extremes are less variable. The positive shape parameter across all three states suggests a heavy-tailed distribution, meaning extreme wind speeds could reach high values. Notably, Texas has a significantly higher shape parameter, implying that the potential for extremely high wind speeds is greater in Texas than in Louisiana and Mississippi. Louisiana's near-zero shape parameter indicates a more moderate extreme tail, meaning less likelihood of very extreme wind speeds compared to Texas and Mississippi.

Texas not only experiences higher thresholds for extreme wind speeds but also shows greater variability and a higher likelihood of extreme wind events. In contrast, Louisiana's extreme wind speeds are less variable and have a more moderate extreme tail, suggesting that very high wind speeds are less common. Mississippi falls between Texas and Louisiana, with moderate variability and a moderate tendency toward extreme wind speeds.

Flood Level:

	TX		LA		MS	
	Estimates	SE	Estimates	SE	Estimates	SE
Location	2.21514	0.11678	2.2146	0.07942	2.2246	0.07845
Scale	0.55906	0.08276	0.3489	0.07326	0.3416	0.07120
Shape	0.05922	0.10169	0.4514	0.22500	0.4223	0.20762

Table 4.3: Flood Level GEV Parameters and Standard Errors

From *Table 4.3*, the near-identical Location parameters across the three states suggest that the threshold for extreme flood levels is very similar in Texas, Louisiana, and Mississippi. This implies that these states experience similar base levels of flooding before the events are categorised as extreme. The higher scale parameter in Texas suggests greater variability in flood levels, meaning that extreme flooding events in Texas have a wider range of severity than in Louisiana and Mississippi. In contrast, the lower scale parameters in Louisiana and Mississippi indicate that the extreme flood levels in these states are more consistent and less variable. The positive shape parameters across all states suggest that the flood level distribution has a heavy tail, meaning there is a potential for very extreme flood events. Louisiana, with the highest shape parameter, indicates the greatest potential for extreme flooding, followed closely by Mississippi. Texas has a much lower shape parameter, suggesting that while there is still potential for extreme flood levels, it is less pronounced compared to the other two states.

The positive shape parameters across all states suggest that the flood level distribution has a heavy tail, meaning there is a potential for very extreme flood events. Louisiana, with the highest shape parameter, indicates the greatest potential for extreme flooding, followed closely by Mississippi. Texas has a much lower shape parameter, suggesting that while there is still potential for extreme flood levels, it is less pronounced compared to the other two states.

4.1.1.2 GEV Diagnostic Plots

Rainfall

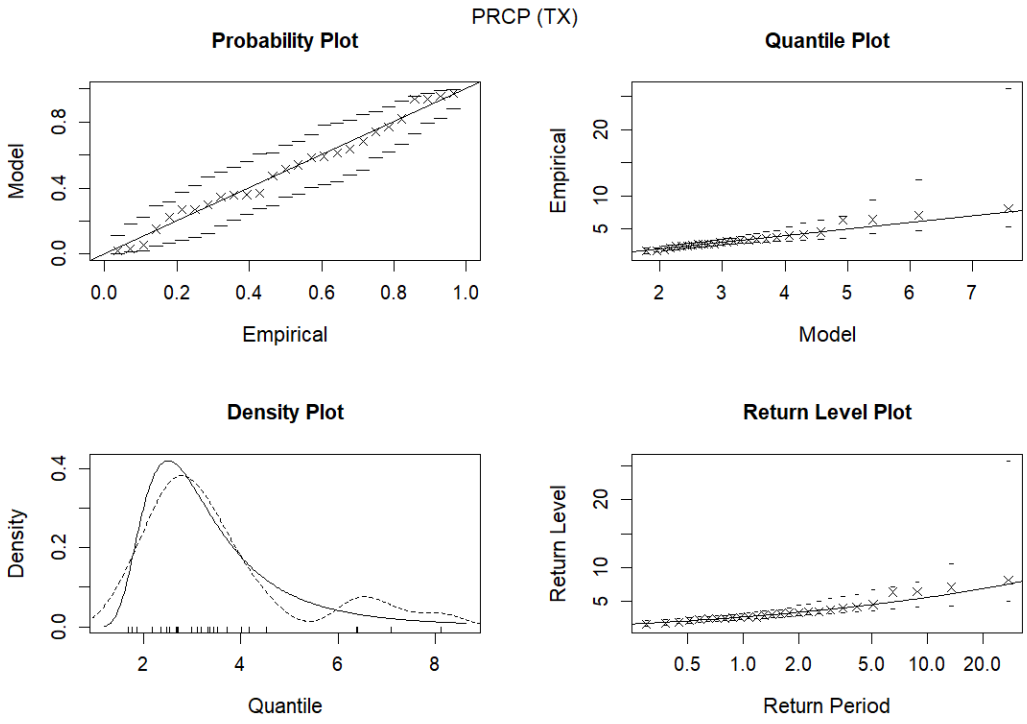

Figure 4.1: GEV Distribution Diagnostic Plots: Rainfall Level in Texas

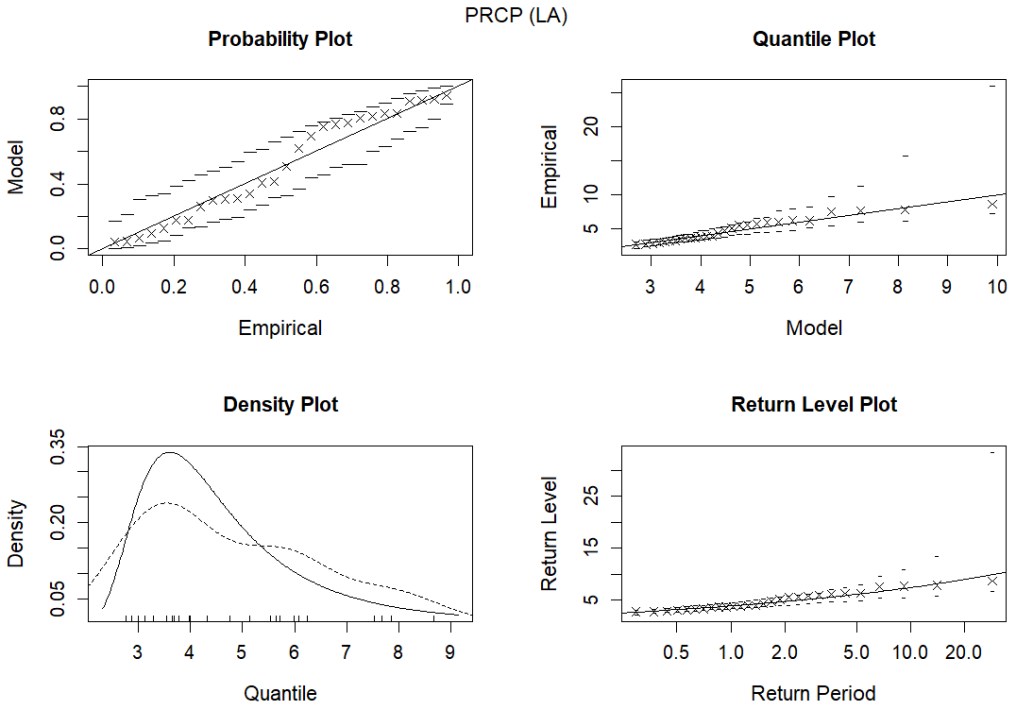

Figure 4.2: GEV Distribution Diagnostic Plots: Rainfall Level in Louisiana

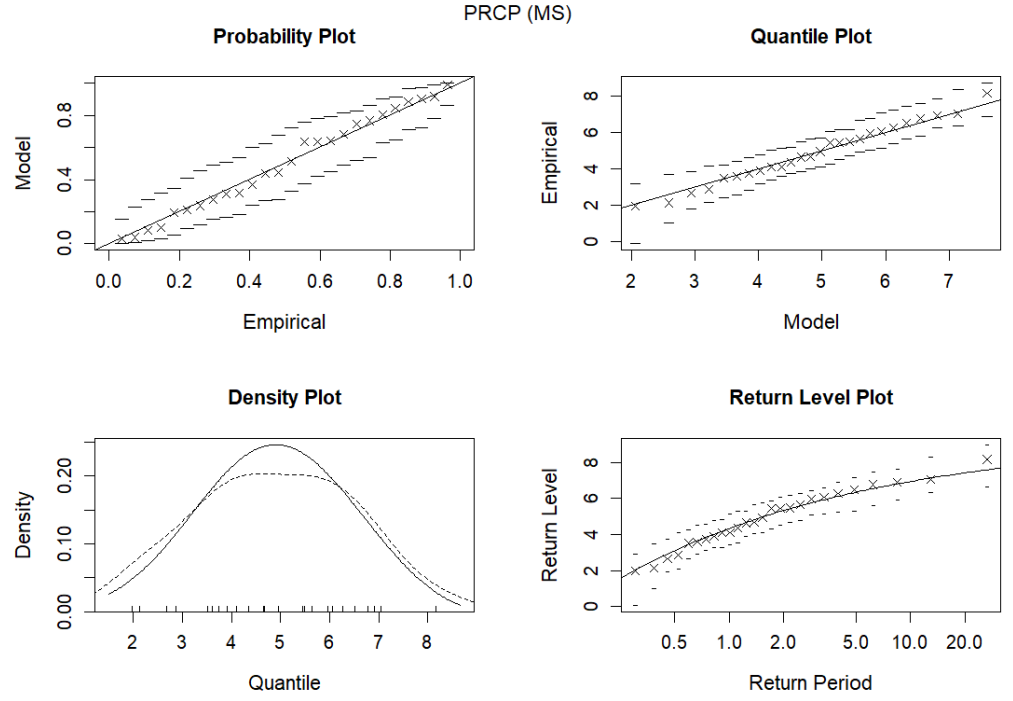

Figure 4.3: GEV Distribution Diagnostic Plots: Rainfall Level in Mississippi

The GEV diagnostic plots (*Figure 4.1, Figure 4.2, Figure 4.3*) suggest that the GEV model is generally a good fit for the precipitation data in Texas, Louisiana, and Mississippi, capturing most of the distribution well. However, there are some areas of concern, particularly at the extreme upper quantiles and higher return periods, where the model underestimates the severity of extreme precipitation events. When making decisions using these GEV models, we should consider the potential limitation brought by the underestimation.

Wind Speed

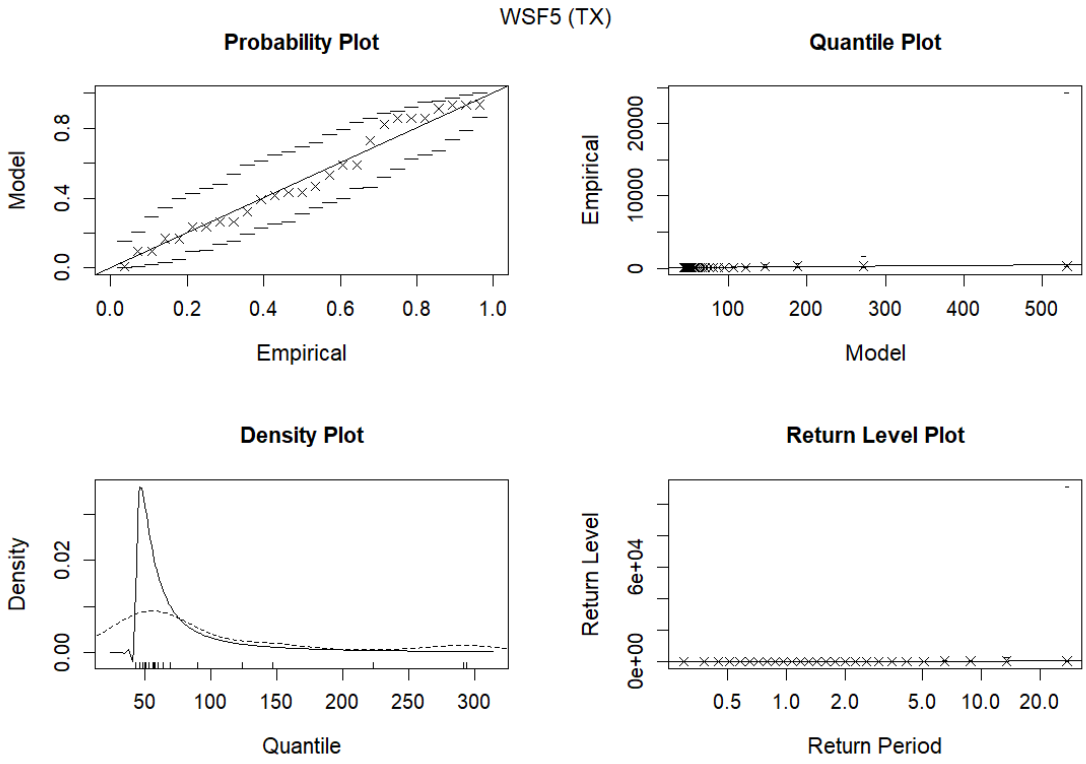

Figure 4.4: GEV Distribution Diagnostic Plots: Wind Speed in Texas

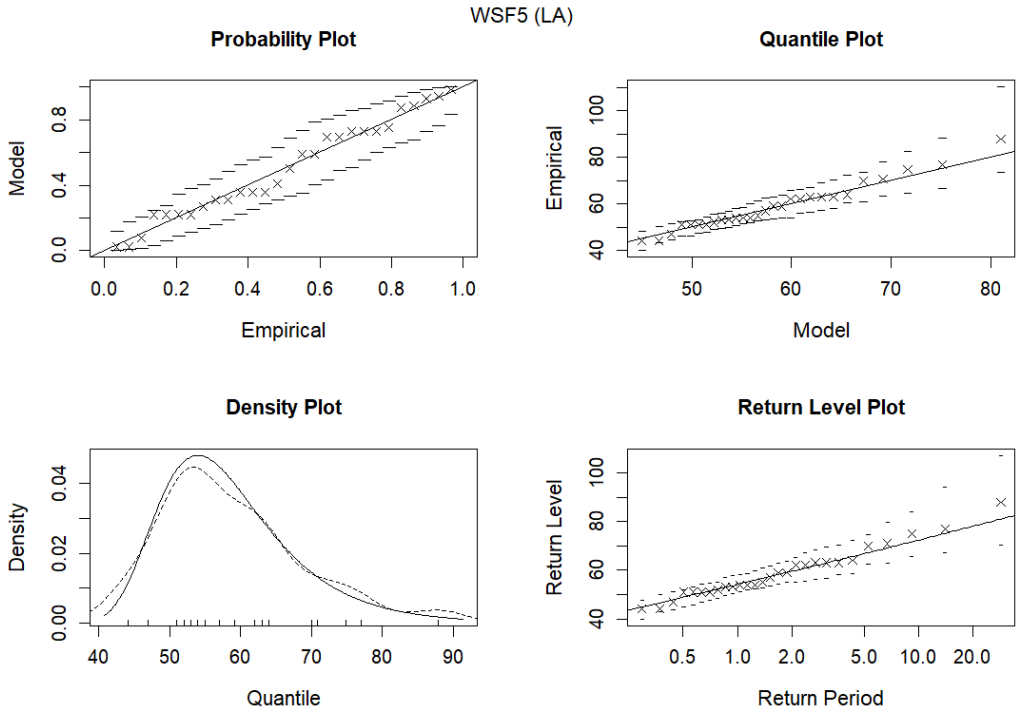

Figure 4.5: GEV Distribution Diagnostic Plots: Wind Speed in Louisiana

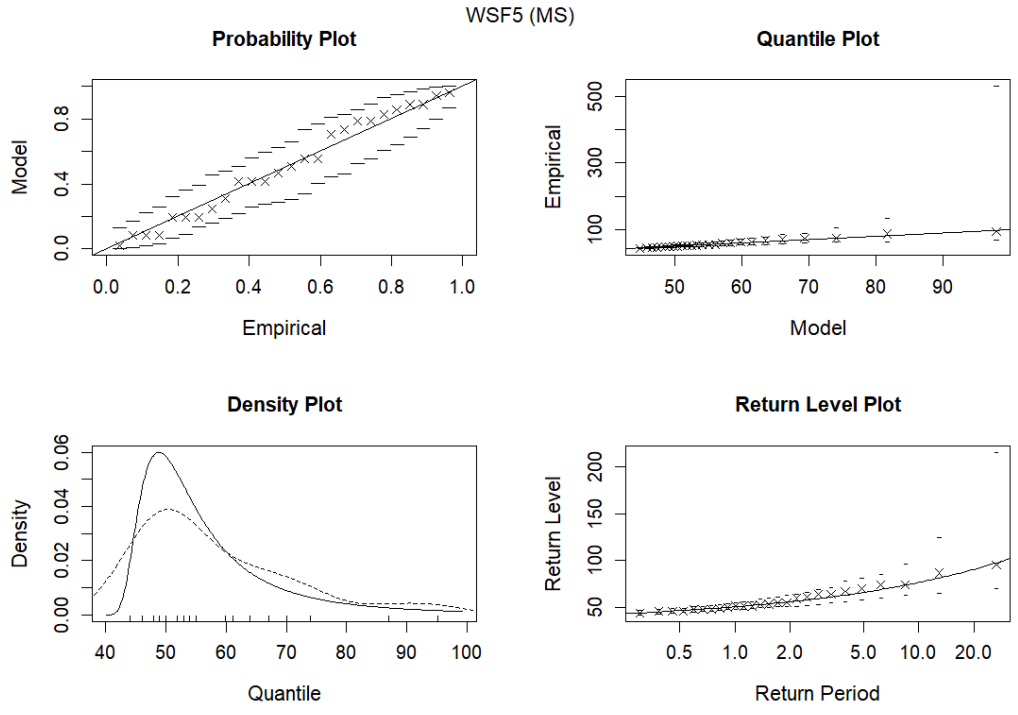

Figure 4.6: GEV Distribution Diagnostic Plots: Wind Speed in Mississippi

The GEV diagnostic plots (*Figure 4.4, Figure 4.5, Figure 4.6*) for wind speed suggest that the GEV model reasonably fits most of the wind speed distribution in Texas, Louisiana, and Mississippi. The model tends to overestimate the windspeed in the lower quantiles, particularly in Texas. However, there are concerns about the higher quantiles and longer return periods, where the model slightly underestimates extreme wind events. This underestimation is more pronounced in Texas and Mississippi, where the model's fit deviates significantly at the upper extremes.

Flood Level

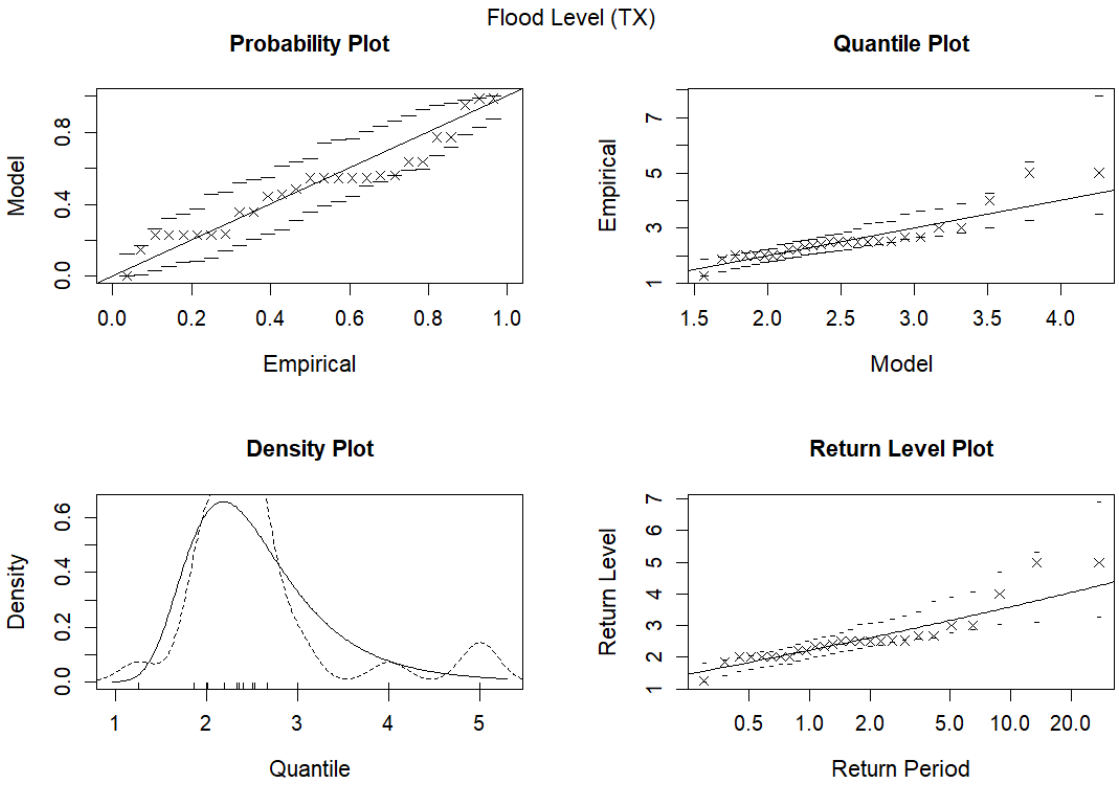

Figure 4.7: GEV Distribution Diagnostic Plots: Flood Level in Texas

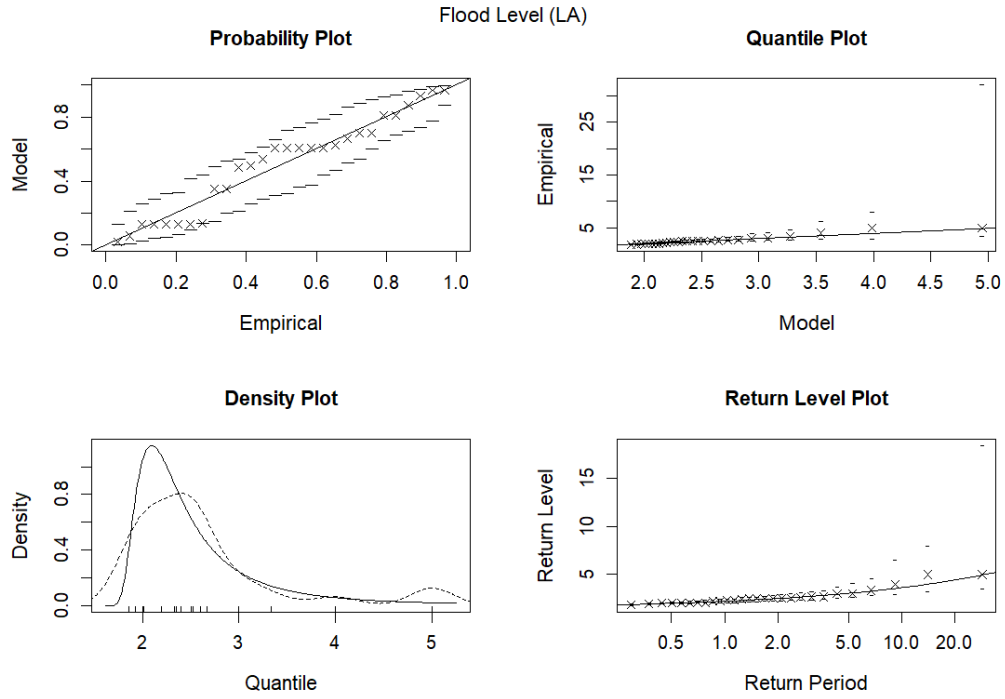

Figure 4.8: GEV Distribution Diagnostic Plots: Flood Level in Louisiana

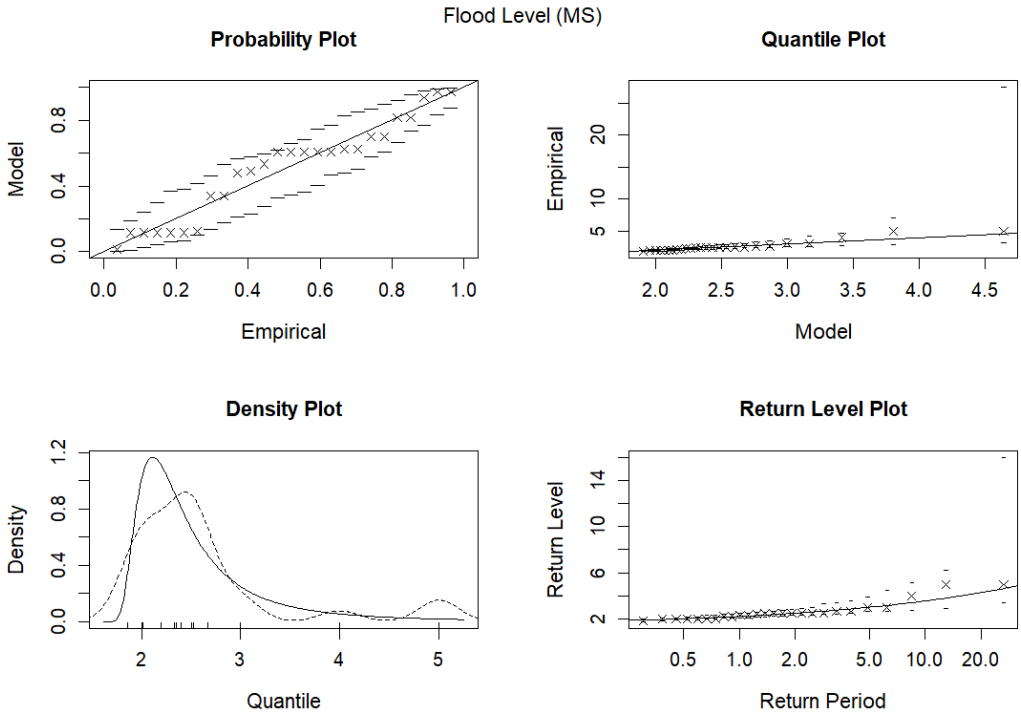

Figure 4.9: GEV Distribution Diagnostic Plots: Flood Level in Mississippi

The GEV diagnostic plots (*Figure 4.7, Figure 4.8, Figure 4.9*) for flood levels in Texas, Louisiana, and Mississippi indicate that while the GEV model generally fits the data well, there are some concerns regarding its ability to predict the most extreme flood levels accurately. Specifically, the model underestimates extreme events' severity, particularly at higher quantiles and longer return periods. This underestimation is consistent across all three states, though it appears slightly more pronounced in Texas.

4.1.1.3 Return Level (GEV)

	TX	LA	MS
Rainfall Level (in)	8.974	11.508	7.919
Wind Speed (mph)	952.352	85.665	116.690
Flood Level (in)	4.669	5.940	5.619

Table 4.4: 50-Year Return Levels for Rainfall, Wind Speed, and Flood Levels (GEV)

The 50-year return level shown in *Table 5.4* suggests that Louisiana is at the greatest risk for extreme rainfall and flood events, with the highest estimated values in these categories (11.508 inches rainfall and 5.940 inches level of flood). While less vulnerable to extreme rainfall, Mississippi shows a considerable risk from both wind and flood events. Despite showing a potential outlier in wind speed, Texas also faces significant risks.

4.1.1.4 Profile Log-likelihood Plots (GEV)

Rainfall

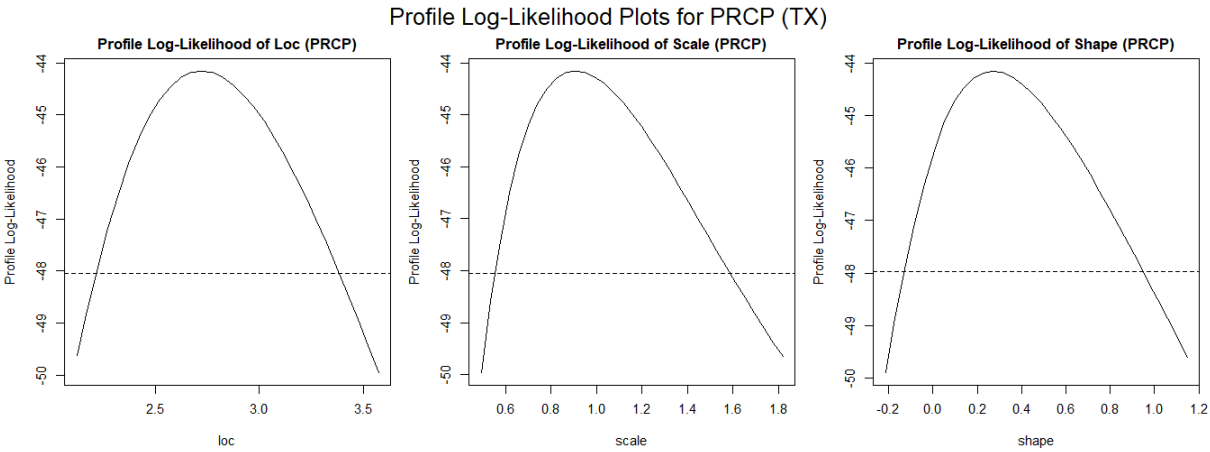

Figure 4.10: Profile Log-Likelihood Plots: Rainfall Level in Texas (GEV)

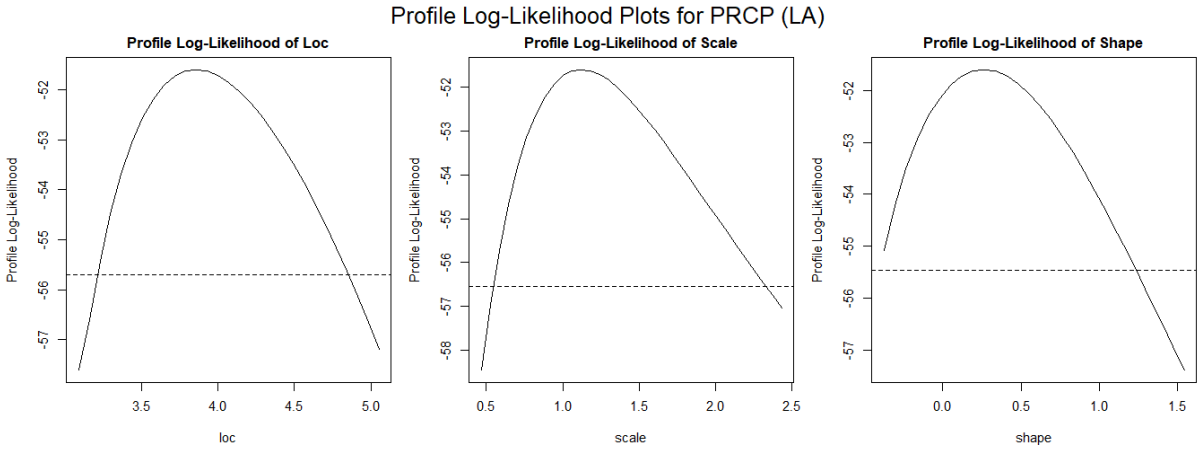

Figure 4.11: Profile Log-Likelihood Plots: Rainfall Level in Louisiana (GEV)

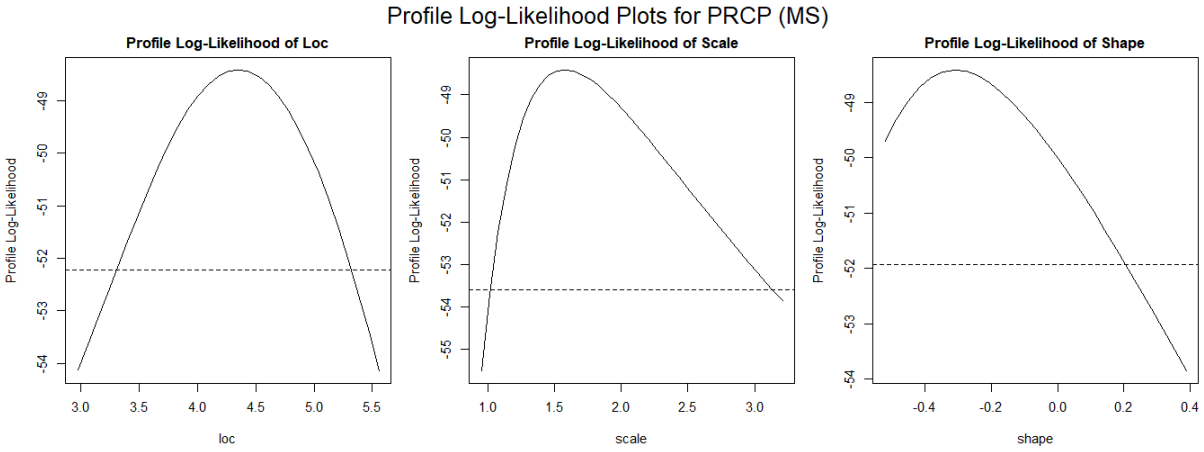

Figure 4.12: Profile Log-Likelihood Plots: Rainfall Level in Mississippi (GEV)

Confident Interval

	TX		LA		MS	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Location	2.329643	3.115107	3.311988	4.410961	3.67409	5.025187
Scale	0.5818325	1.2316714	0.6543816	1.5896617	1.08831	2.07056
Shape	-0.076071	0.61321172	-0.293719	0.8206814	-0.5851	-0.03687

Table 4.5: Confidence Intervals of Rainfall GEV Parameters

Wind Speed

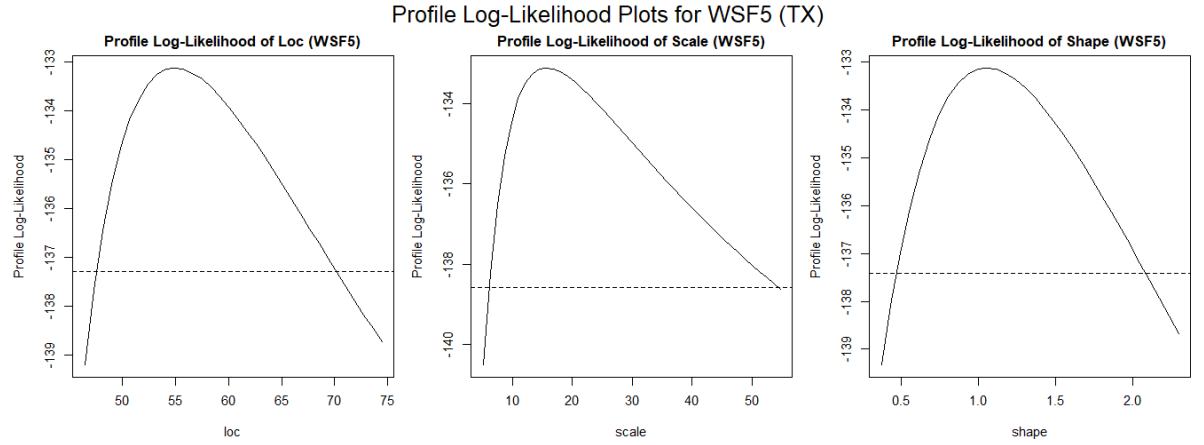

Figure 4.13 Profile Log-Likelihood Plots: Windspeed in Texas (GEV)

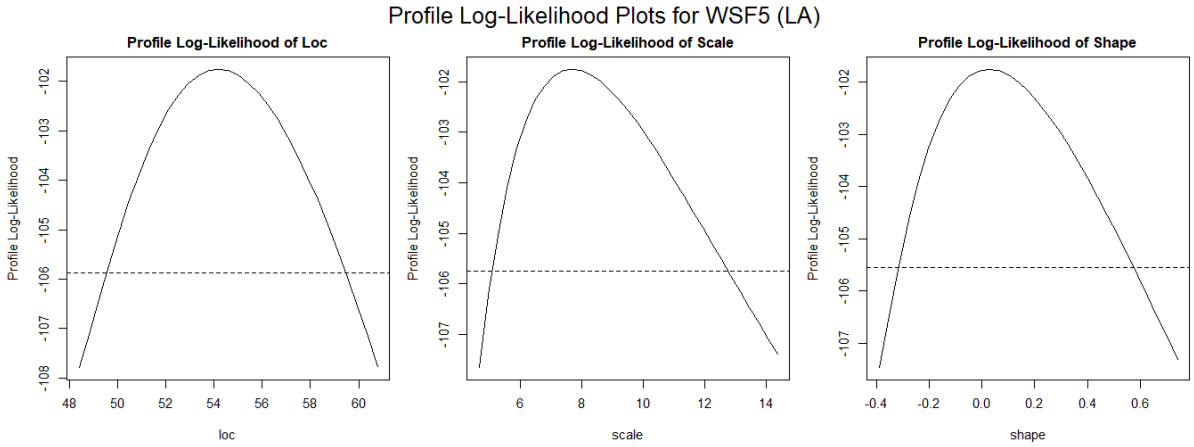

Figure 4.14 Profile Log-Likelihood Plots: Windspeed in Louisiana (GEV)

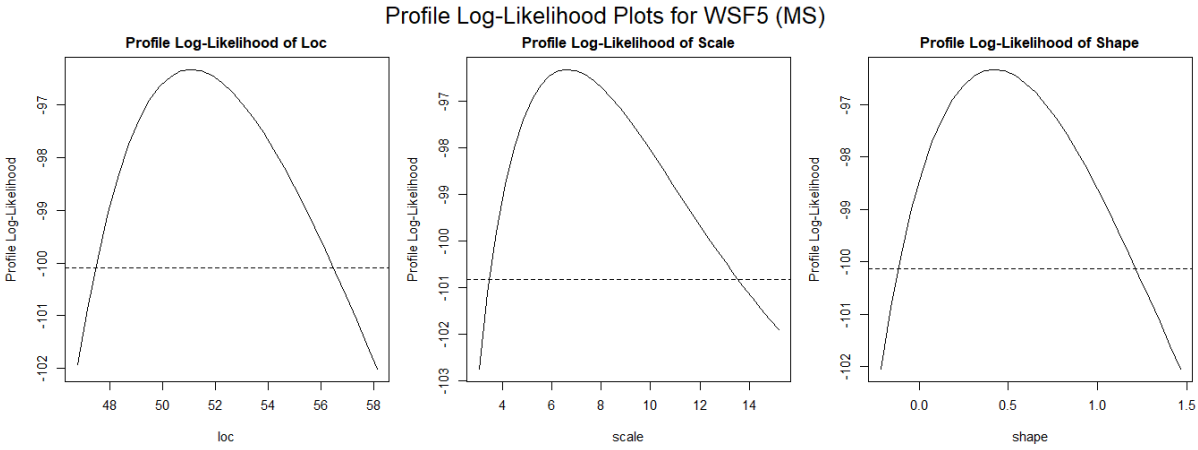

Figure 4.15 Profile Log-Likelihood Plots: Windspeed in Mississippi (GEV)

Confident Interval

	TX		LA		MS	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Location	48.27947	61.58541	50.94279	57.41153	48.00850	54.14366
Scale	6.454266	24.988010	5.301009	10.065210	3.832706	9.424710
Shape	0.5670855	1.5415552	-0.27073	0.3206511	-0.03397	0.87559

Table 4.6: Confidence Intervals of Windspeed GEV Parameters

Flood Level

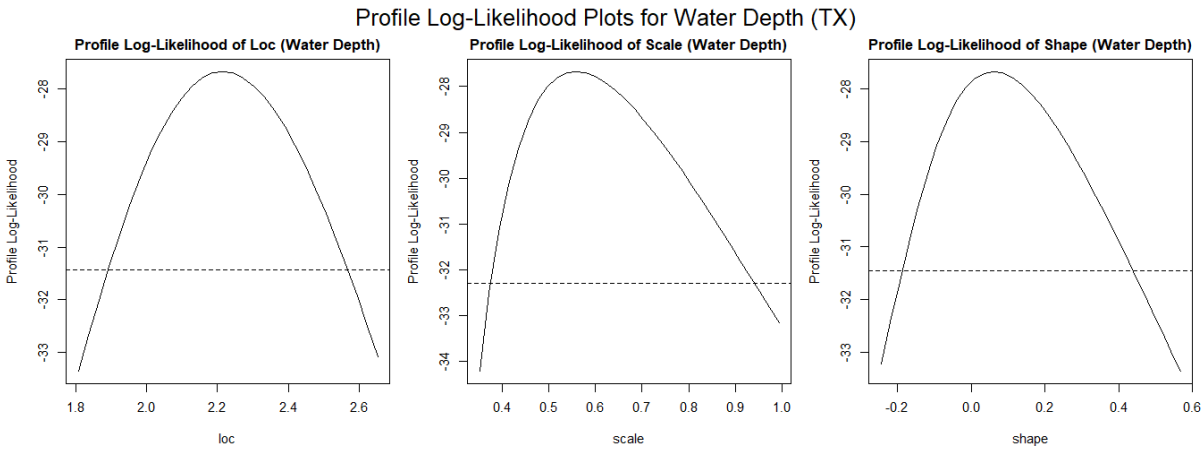

Figure 4.16 Profile Log-Likelihood Plots: Flood Level in Texas (GEV)

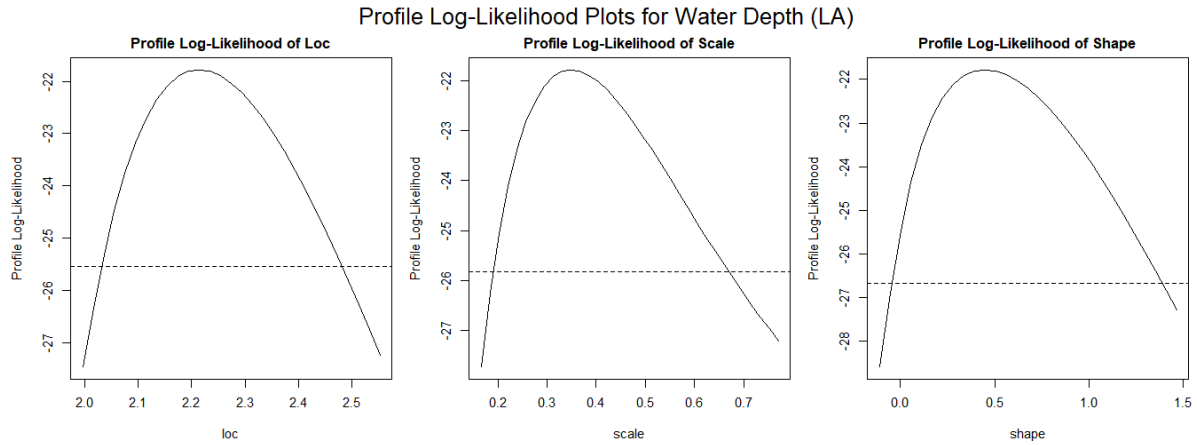

Figure 4.17 Profile Log-Likelihood Plots: Flood Level in Louisiana (GEV)

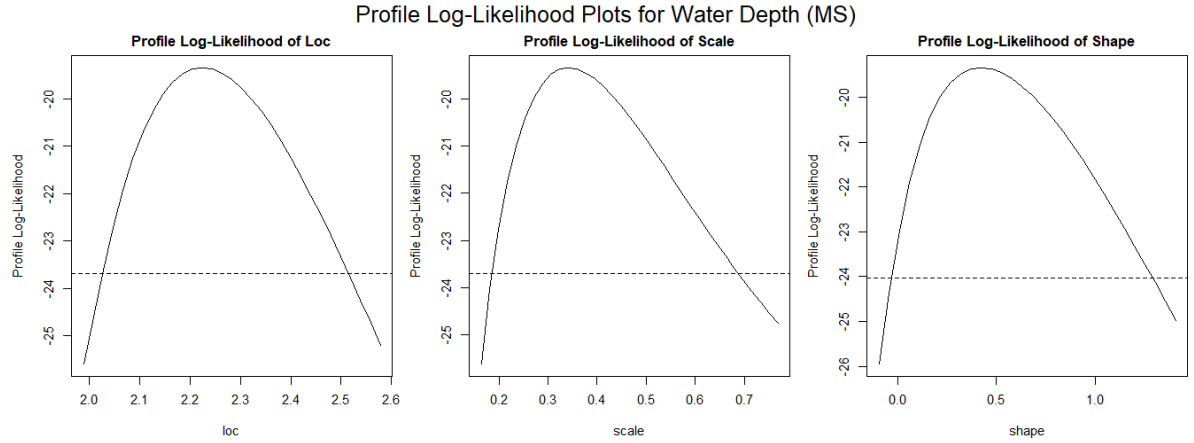

Figure 4.18 Profile Log-Likelihood Plots: Flood Level in Mississippi (GEV)

Confident Interval

	TX		LA		MS	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Location	1.986259	2.4440	2.058941	2.370246	2.070816	2.378353
Scale	0.39685	0.7213	0.2052971	0.4924757	0.2020824	0.4811669
Shape	-0.14009	0.2585	0.01039515	0.89238266	0.0153925	0.8292368

Table 4.7: Confidence Intervals of Flood Level GEV Parameters

The profile log-likelihood plots show the likelihood of different values for each parameter (location, scale, and shape) while holding the other parameters fixed at their estimated values. The peak of each curve represents the maximum likelihood estimate for that parameter, and the curve's shape indicates the estimate's precision. Steeper slopes and narrower peaks generally indicate that the parameter estimates are more precise, meaning the model fits the data better around those parameter values. Flatter slopes and broader peaks suggest more uncertainty in the parameter estimates, indicating a more moderate fit.

The lack of steep slopes and narrow peaks in the profile log-likelihood plots (*Figure 4.10, Figure 4.11, Figure 4.12, Figure 4.13, Figure 4.14, Figure 4.15, Figure 4.16, Figure 4.17, Figure 4.18*) suggests that while the model is likely capturing the general structure of the data, it may not be capturing all nuances with high precision.

From *Table 4.5, Table 4.6* and *Table 4.7*, narrower confidence intervals (e.g., location parameters) suggest greater certainty in parameter estimates, while wider intervals (e.g., scale and shape parameters) reflect greater uncertainty. The shape parameters indicate whether the model predicts the possibility of very extreme events (heavy tails) or if such extremes are less likely (bounded tails). While there are different signs between the lower bond (2.5%) and the upper bond (97.5%), we need further consideration to examine the value of shape parameters to decide the tail behaviours.

5.1.1.5 Likelihood Ratio Test (GEV)

Deviance difference	TX	LA	MS
Rainfall Level	3.16	0.94	3.18
Windspeed	31.57	0.03	4.23
Flood Level	0.38	7.45	7.66

Table 4.8: Likelihood Ratio Test Results for Rainfall, Windspeed, and Flood Levels

We conducted a Likelihood Ratio Test (LRT) for rainfall, windspeed, and flood levels across Texas, Louisiana, and Mississippi to evaluate whether the more complex GEV model provides a significantly better fit compared to a simpler alternative. The results are summarised in *Table 4.8*.

Rainfall Level:

In Texas and Mississippi, the LRT showed deviance differences of 3.16 and 3.18, respectively, suggesting that the more complex GEV model provides a better fit for extreme rainfall events in these states. In Louisiana, however, the deviance difference was only 0.94, indicating that the simpler model might be acceptable for this state.

Windspeed:

The deviance difference for windspeed in Texas was 31.57, strongly favouring the more complex model. In Mississippi, the moderate deviance difference of 4.23 suggests some improvement in model fit. In contrast, Louisiana's deviance difference was negligible (0.03), implying that the simpler model also works for windspeed data in this state.

Flood Level:

For flood levels, Louisiana and Mississippi showed substantial deviance differences (7.45 and 7.66, respectively), indicating that the more complex model significantly improves the fit. Texas, with a deviance difference of only 0.38, shows minimal improvement, suggesting that the simpler model might be sufficient.

The LRT results indicate that the more complex GEV model is generally justified for predicting windspeed in Texas and flood levels in Louisiana and Mississippi. However, for rainfall in Louisiana and flood levels in Texas, the simpler model might be adequate, suggesting that model complexity should be tailored to the specific variable and location.

4.1.1.6 Goodness of Fit

Rainfall Level:

	TX	LA	MS
Kolmogorov-Smirnov test	0.9801	0.54	0.9542
Cramer-von Mises test	0.9709	0.66	0.9956

Table 4.9: Goodness of Fit Results for Rainfall Level GEV Model

Windspeed:

	TX	LA	MS
Kolmogorov-Smirnov test	0.834	0.8699	0.9765
Cramer-von Mises test	0.8254	0.88	0.9507

Table 4.10: Goodness of Fit Results for Windspeed GEV Model

Flood Level:

	TX	LA	MS
Kolmogorov-Smirnov test	0.3373	0.5714	0.6434
Cramer-von Mises test	0.3313	0.5603	0.5018

Table 4.11: Goodness of Fit Results for Flood Level GEV Model

From *Table 4.9*, *Table 4.10* and *Table 4.11*, across all variables and states, the test results for both *Kolmogorov-Smirnov (K-S)* and *Cramer-von Mises (C-vM)* tests indicate that we failed to reject the null hypothesis, suggesting that the data are likely to come from the fitted GEV distributions, confirming that the GEV models are generally appropriate for modelling extreme events across all variables and states.

The GEV models for rainfall and windspeed show particularly strong fits in Texas and Mississippi, with high p-values indicating that the models accurately capture the distribution of extreme events in these states. Undeniably, the results indicate a more moderate fit for the GEV models in Louisiana, particularly for flood and rainfall levels. The lower *p-values* here, while still acceptable, suggest that the model may not capture all aspects of the data perfectly, indicating some limitations in modelling the extreme events in this state.

4.1.2. Peak Over Threshold Approach

4.1.2.1 GPD Threshold Analysis

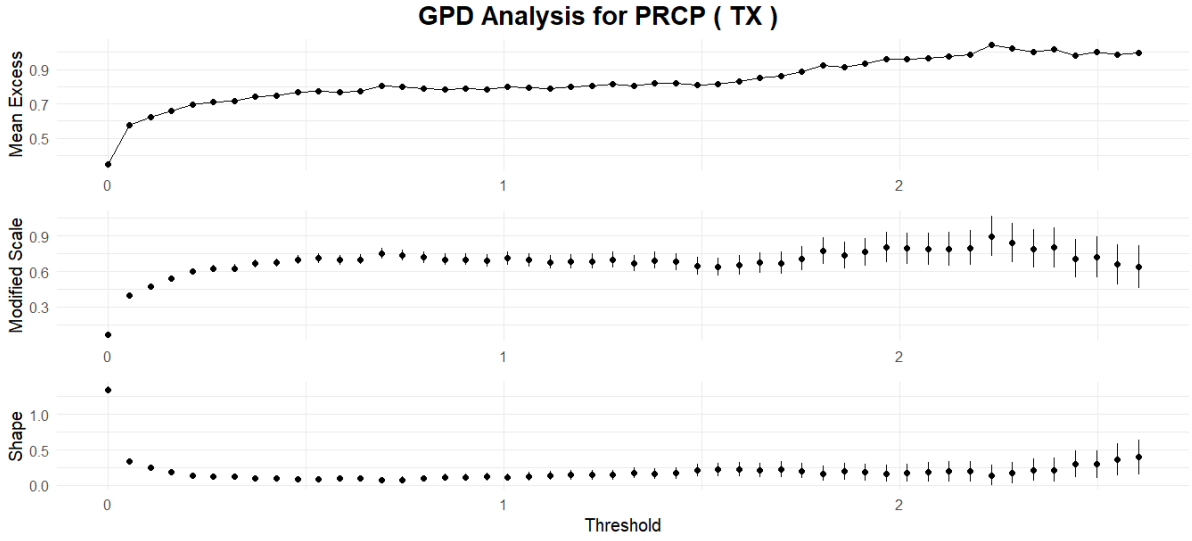

Figure 4.19 GPD Threshold Analysis Plots: Rainfall Level in Texas

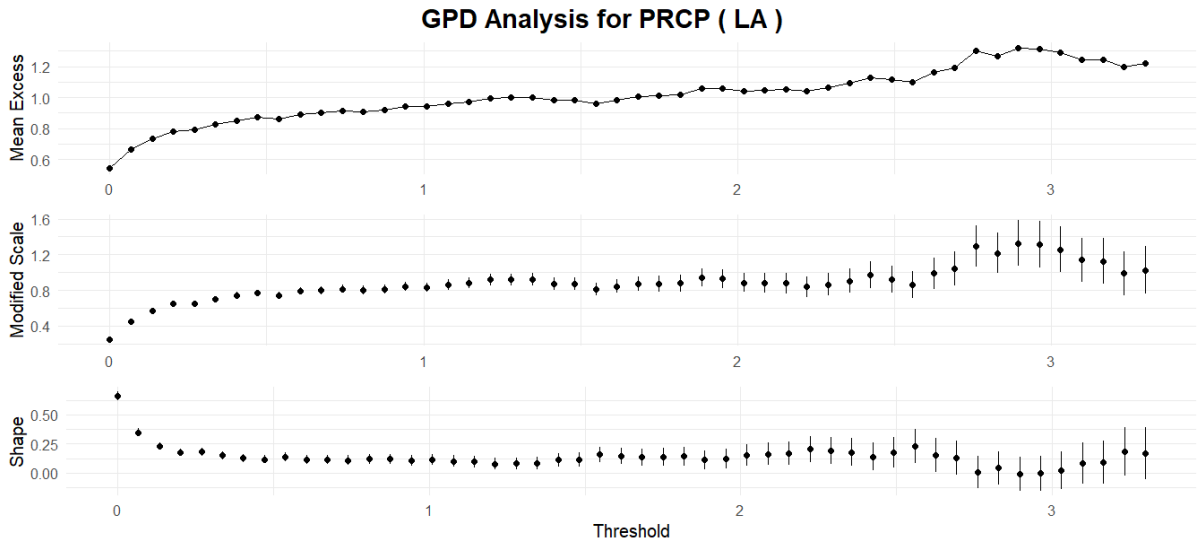

Figure 4.20 GPD Threshold Analysis Plots: Rainfall Level in Louisiana

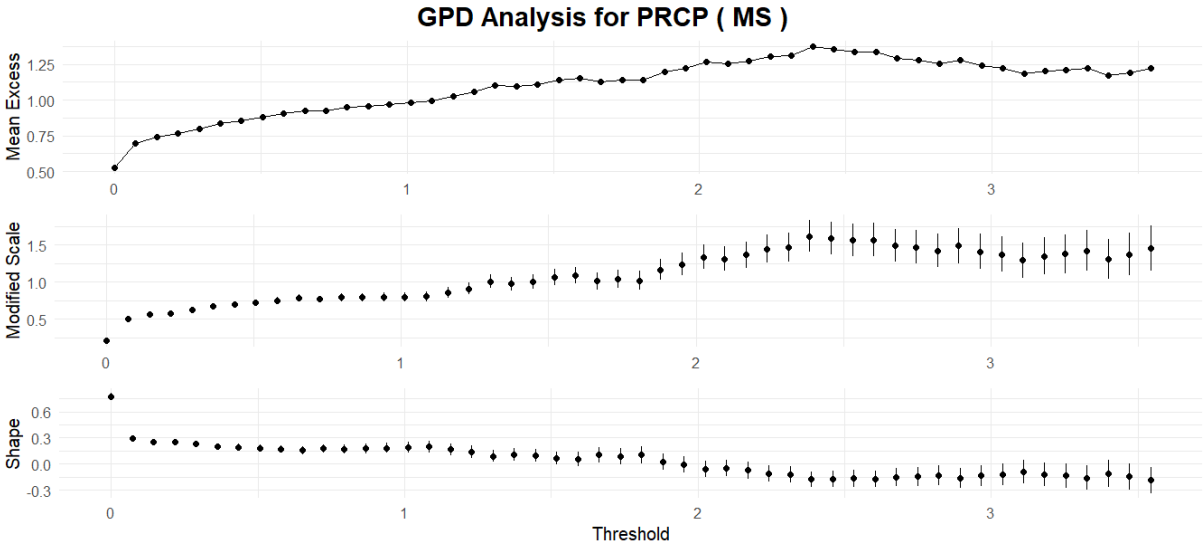

Figure 4.21 GPD Threshold Analysis Plots: Rainfall Level in Mississippi

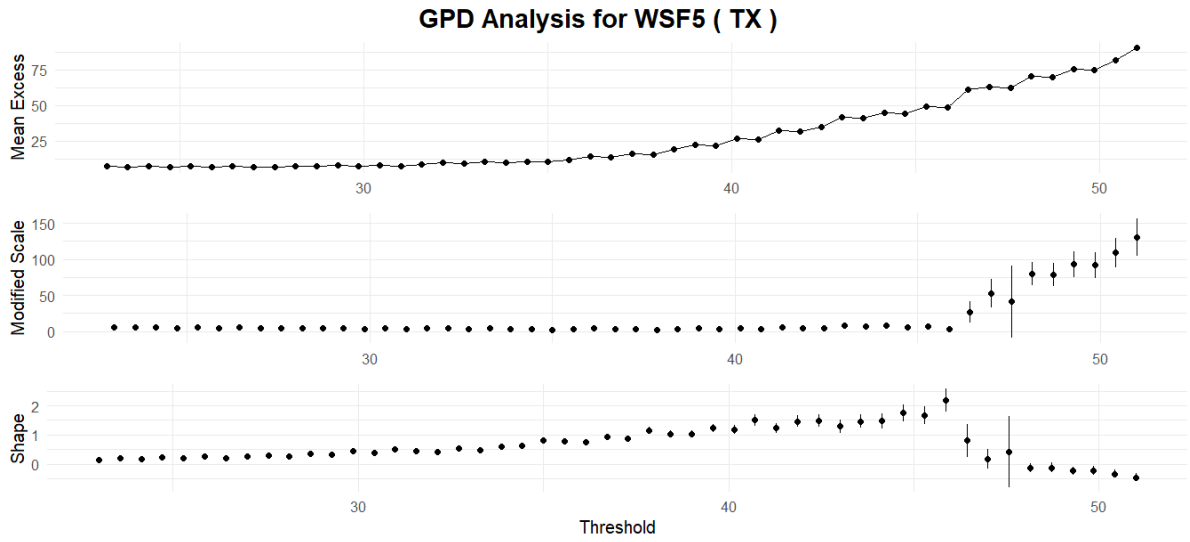

Figure 4.22 GPD Threshold Analysis Plots: Windspeed in Texas

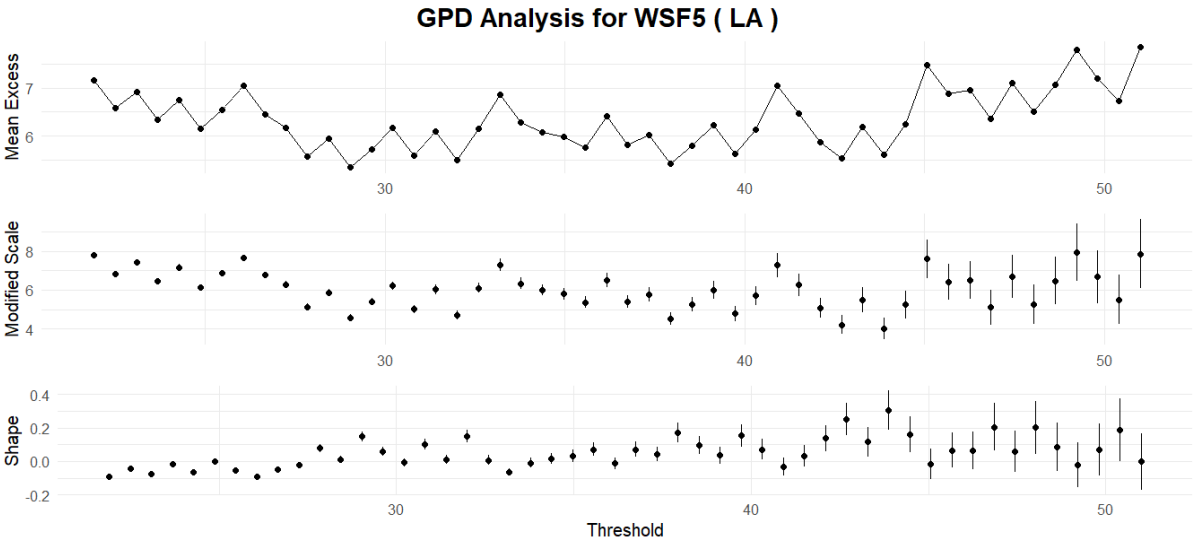

Figure 4.23 GPD Threshold Analysis Plots: Windspeed in Louisiana

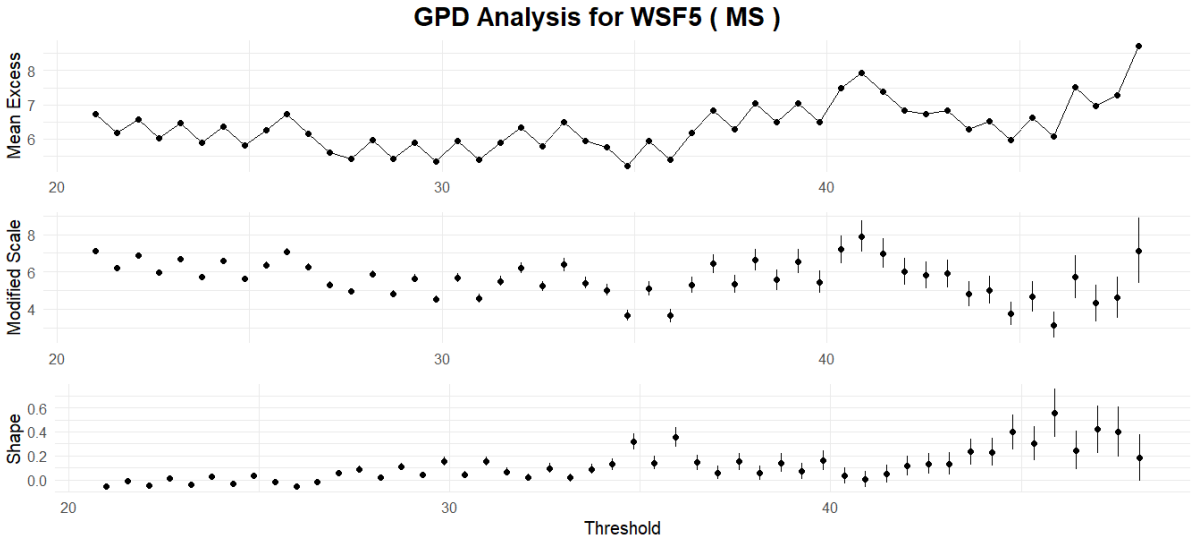

Figure 4.24 GPD Threshold Analysis Plots: Windspeed in Mississippi

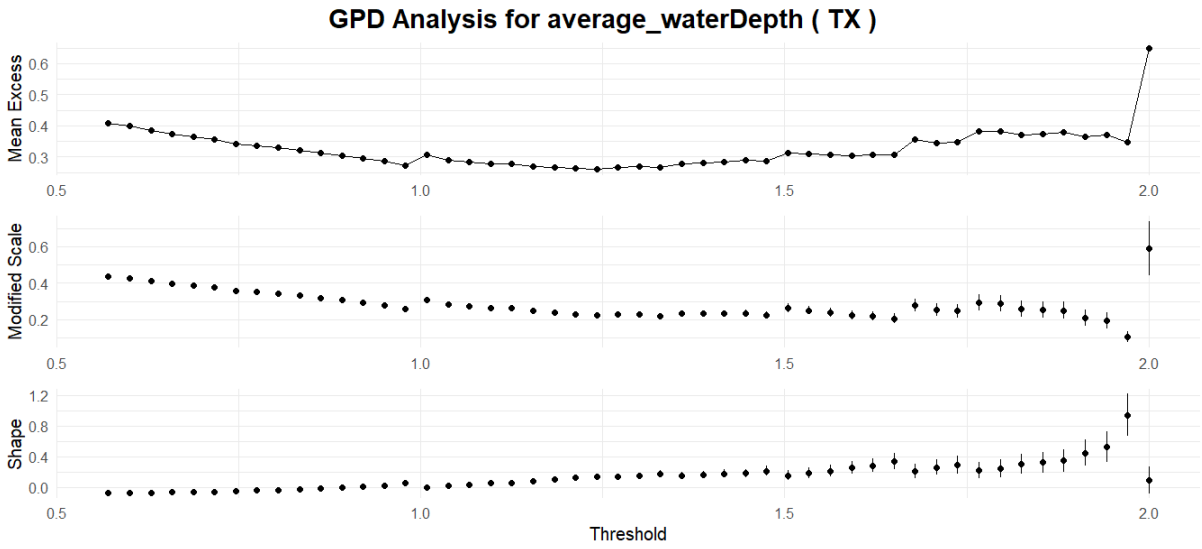

Figure 4.25 GPD Threshold Analysis Plots: Flood Level in Texas

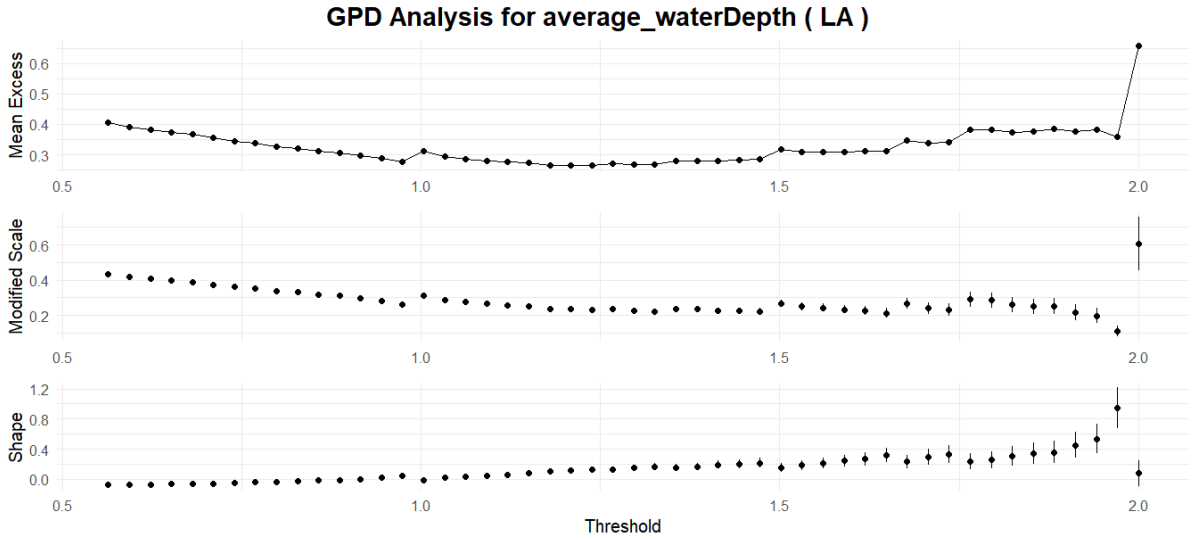

Figure 4.26 GPD Threshold Analysis Plots: Flood Level in Louisiana

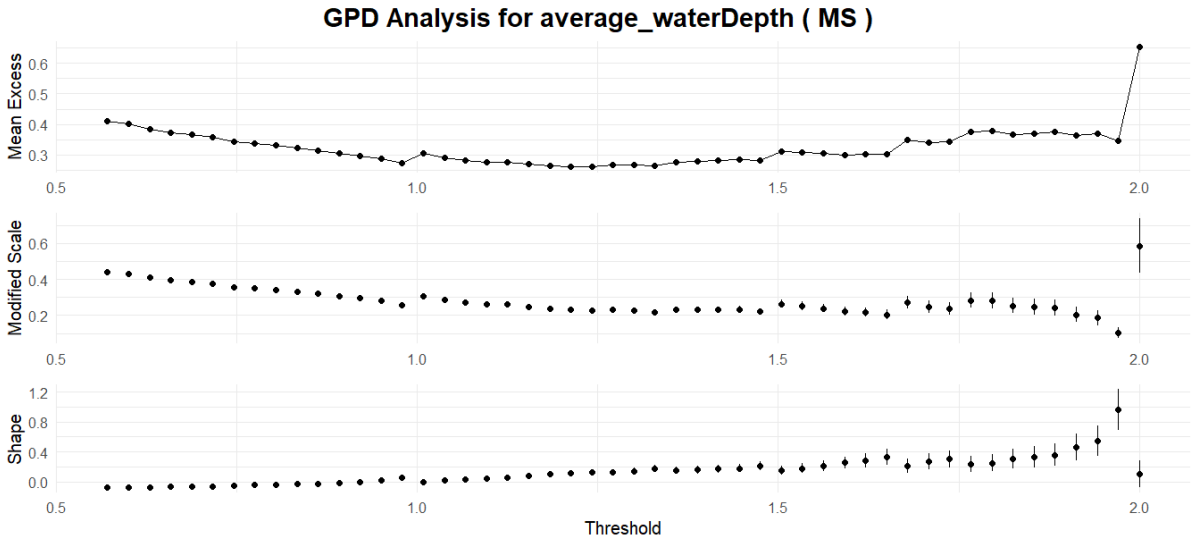

Figure 4.27 GPD Threshold Analysis Plots: Flood Level in Mississippi

Thresholds Selection:

Threshold	TX	LA	MS
Rainfall Level (in)	1.6	1.5	1.3
Windspeed (mph)	45	46	46

Flood Level (m)	1.5	1.3	1.2

Table 4.12: Selected Thresholds for GPD

Thresholds are selected via the trend in threshold analysis plots to become more stable. From *Figure 4.19, Figure 4.20, Figure 4.21, Figure 4.22, Figure 4.23, Figure 4.24, Figure 4.25, Figure 4.26,* and *Figure 4.27*, the selected thresholds are in *Table 4.12*.

5.1.2.2 GPD Parameter Estimation

Rainfall:

TX			LA		MS	
	Estimates	SE	Estimates	SE	Estimates	SE
Scale	0.6699	0.08448	0.8502	0.07235	1.01017	0.08944
Shape	0.2077	0.09885	0.1296	0.06412	0.08633	0.06869

Table 4.13: Rainfall GPD Parameters and Standard Errors

From *Table 4.13*, the increasing scale parameter from Texas to Mississippi suggests that as one moves eastward, there is greater variability in the intensity of extreme rainfall events. Mississippi, in particular, may experience a wider range of extreme rainfall outcomes compared to Texas and Louisiana. The increasing scale parameter from Texas to Mississippi suggests that as one moves eastward, there is greater variability in the intensity of extreme rainfall events. Mississippi, in particular, may experience a wider range of extreme rainfall outcomes compared to Texas and Louisiana.

Mississippi exhibits the greatest variability (as indicated by the scale parameter), while Texas has the most pronounced heavy-tail behaviour (as indicated by the shape parameter), suggesting a higher likelihood of extreme rainfall events in these states compared to Louisiana.

Wind Speed:

	TX		LA		MS	
	Estimates	SE	Estimates	SE	Estimates	SE
Scale	8.532	2.2480	5.6512	0.8734	2.7080	0.6671
Shape	1.530	0.2904	0.1374	0.1210	0.6704	0.2330

Table 4.14: Windspeed GPD Parameters and Standard Errors

From *Table 4.14*, the variation in the scale parameter across the states suggests that Texas experiences the most variable extreme wind speeds, followed by Louisiana and then Mississippi. This indicates that extreme wind events in Texas could be much more severe and unpredictable compared to the other states. The shape parameter results indicate significant differences in the tail behaviour of extreme wind speeds across the states. Texas, with the highest shape parameter, is particularly vulnerable to very extreme wind events, whereas Louisiana's wind speeds tend to be less extreme, with Mississippi falling somewhere in between. Texas shows both high variability and a strong tendency towards extremely high wind speeds, making it particularly susceptible to severe wind events. Louisiana, in contrast, exhibits lower variability and a much lighter tail, indicating fewer and less extreme wind events. Mississippi has moderate characteristics, with some potential for extreme events but less variability compared to Texas.

Flood Level:

	TX		LA		MS	
	Estimates	SE	Estimates	SE	Estimates	SE
Scale	0.2699	0.02288	0.2315	0.01274	0.23883	0.01070
Shape	0.1432	0.05882	0.1411	0.03972	0.09793	0.03083

Table 4.15: Flood Level GPD Parameters and Standard Errors

From *Table 4.15*, the consistency of the scale parameters across the states suggests that the variability of extreme flood levels is relatively similar in Texas, Louisiana, and Mississippi. However, Texas exhibits slightly higher variability (higher scale parameter), which may indicate a broader range of extreme flood outcomes. Texas and Louisiana show a slightly higher propensity for very extreme flood events than Mississippi. The positive shape values across all states highlight the risk of significant outliers, particularly in flood-prone areas. The tail behaviour (as indicated by the shape parameter) suggests that all three states are prone to extreme flood events, with Texas and Louisiana having a slightly higher likelihood of experiencing more extreme events compared to Mississippi.

4.1.2.3 GPD Diagnostic Plots

Rainfall

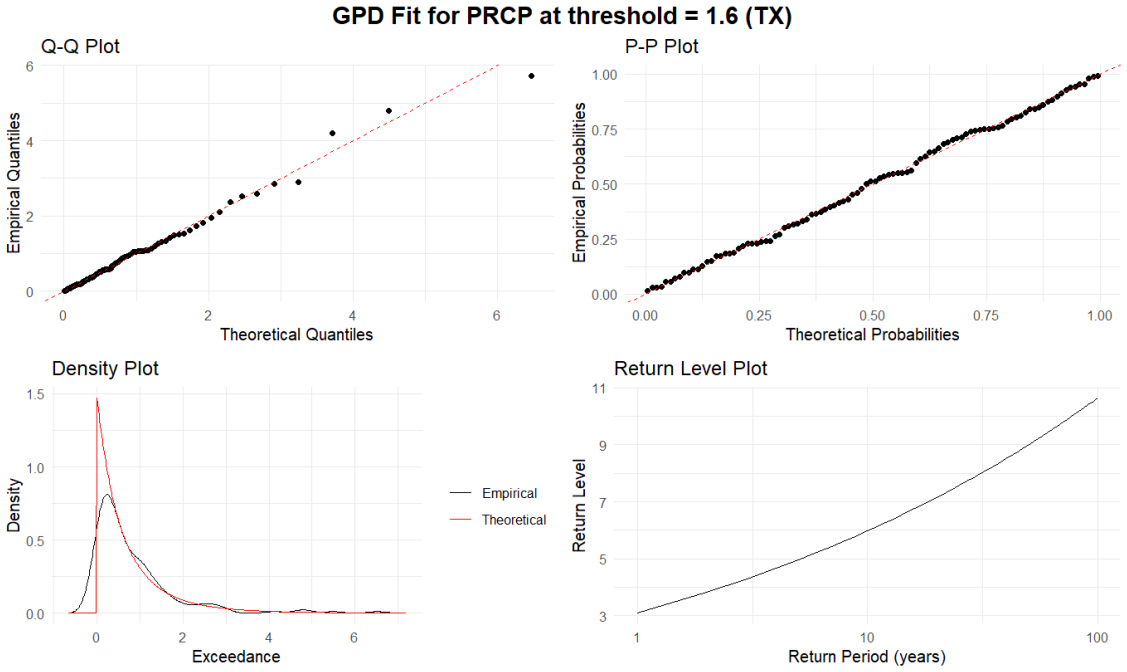

Figure 4.28: GPD Distribution Diagnostic Plots: Rainfall Level in Texas

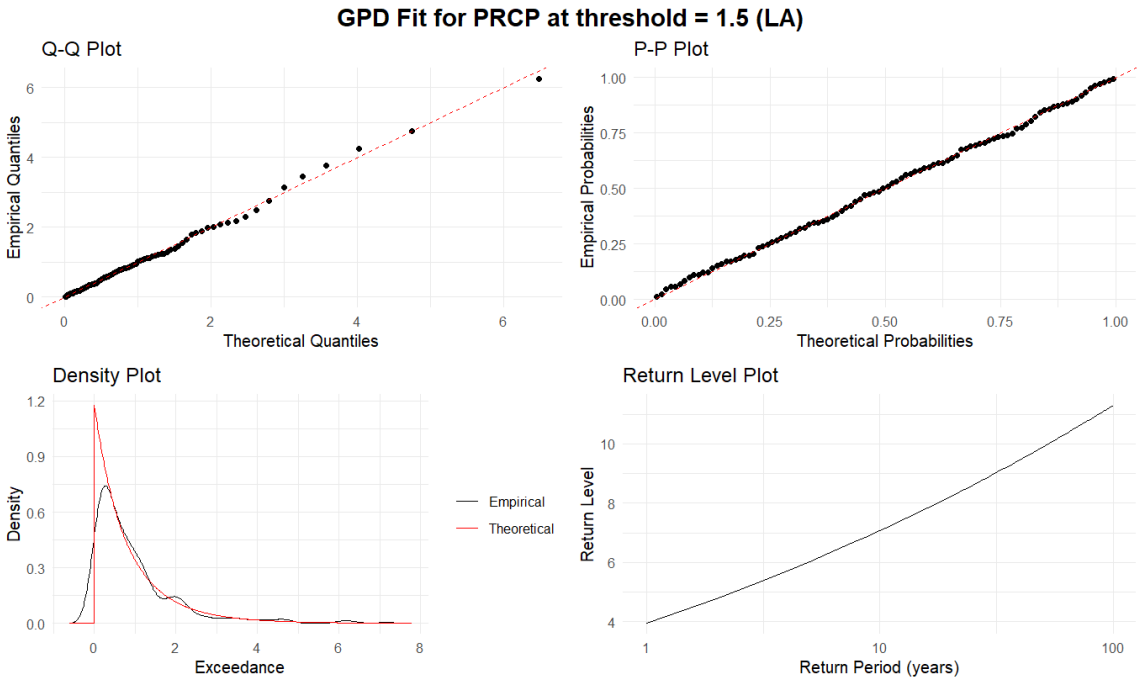

Figure 4.29: GPD Distribution Diagnostic Plots: Rainfall Level in Louisiana

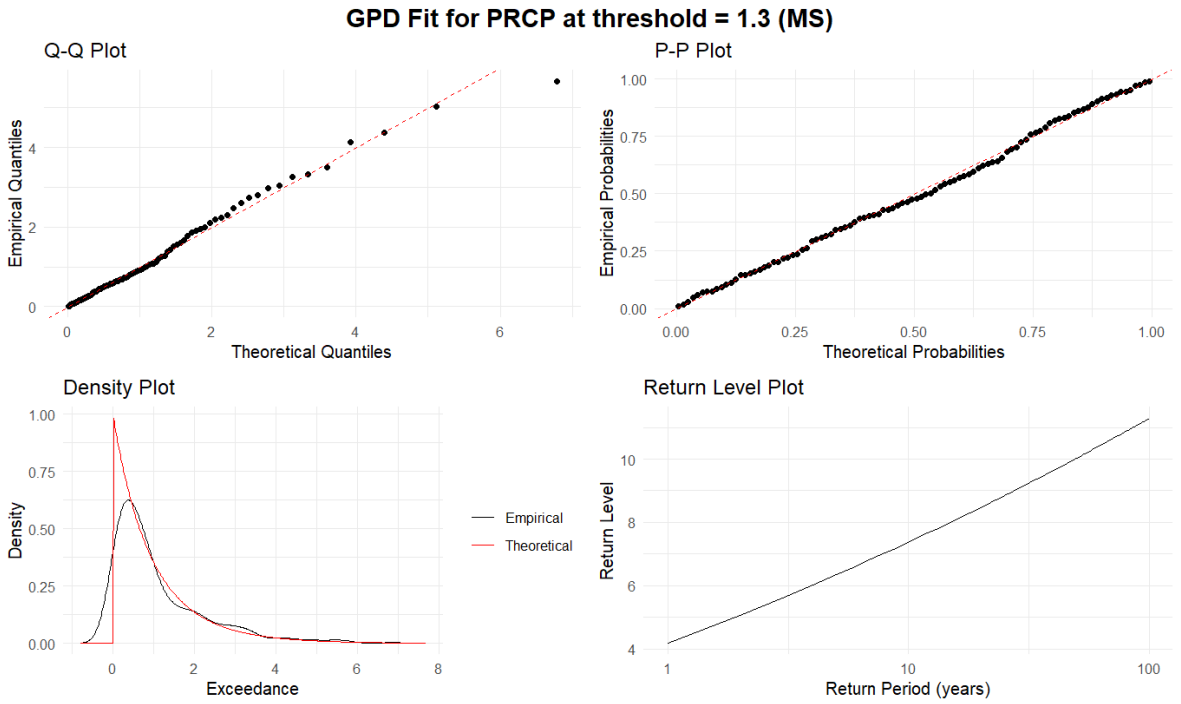

Figure 4.30: GPD Distribution Diagnostic Plots: Rainfall Level in Mississippi

From the *P-P* and *Q-Q* plots in *Figure 4.28, Figure 4.29* and *Figure 30*, the GPD model fits the rainfall level data well. However, in the density plot, there are areas where it underestimates the frequency of certain exceedance values.

Wind Speed

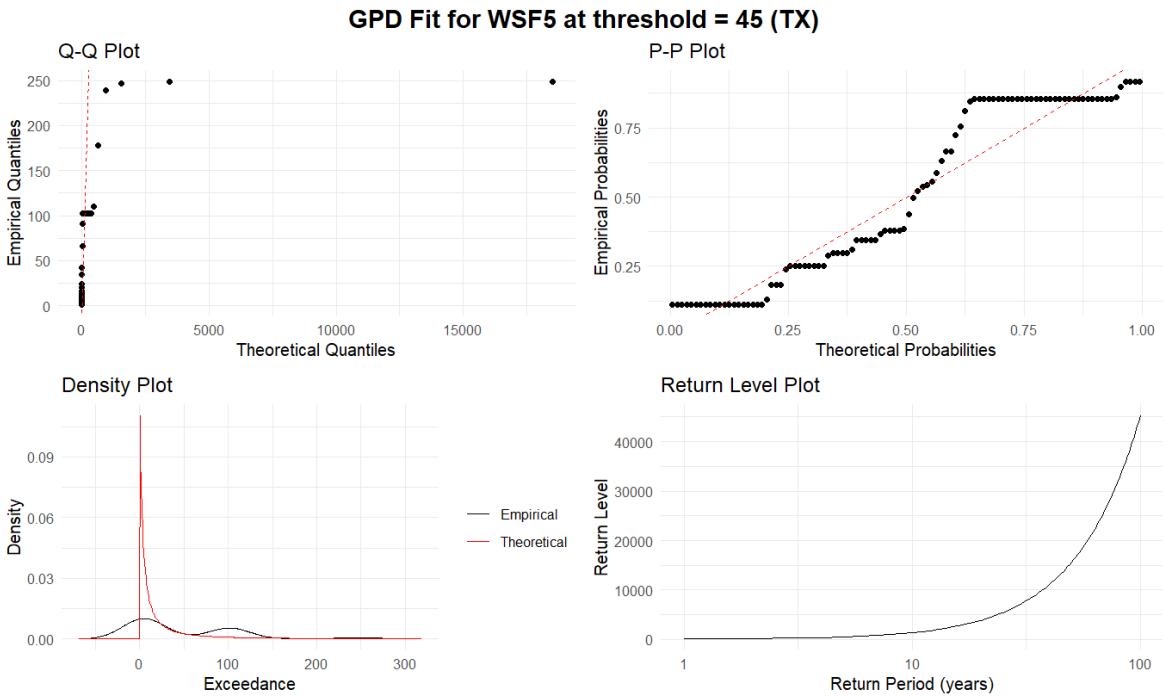

Figure 4.31 GPD Distribution Diagnostic Plots: Wind Speed in Texas

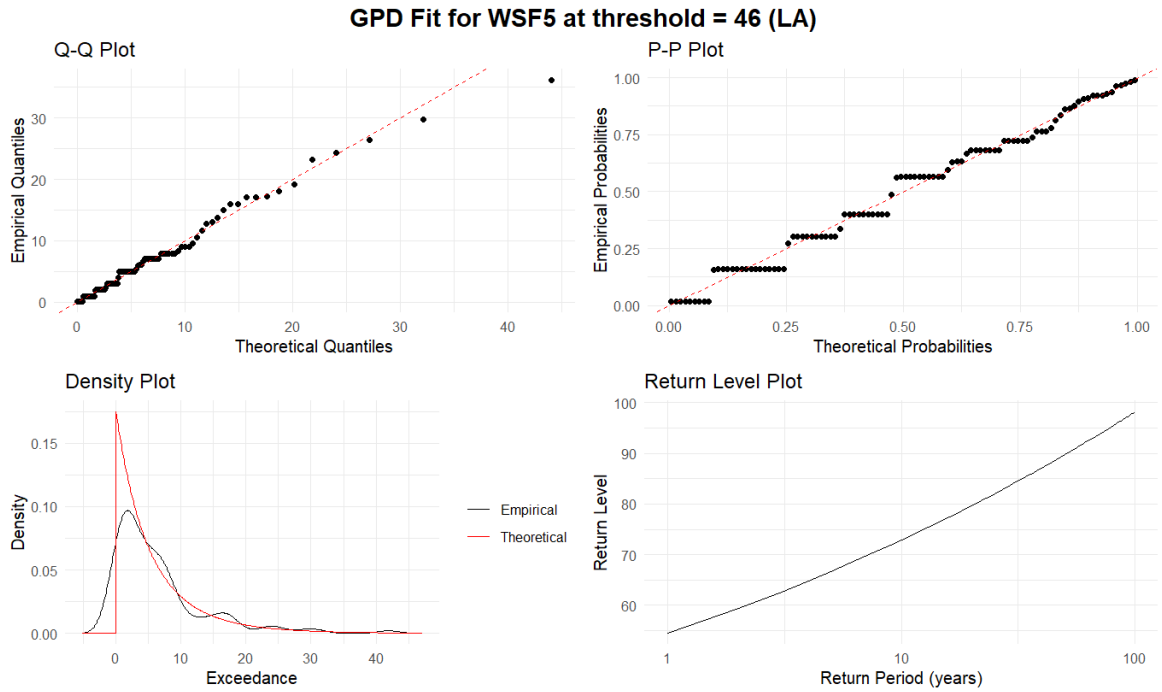

Figure 4.32 GPD Distribution Diagnostic Plots: Wind Speed in Louisiana

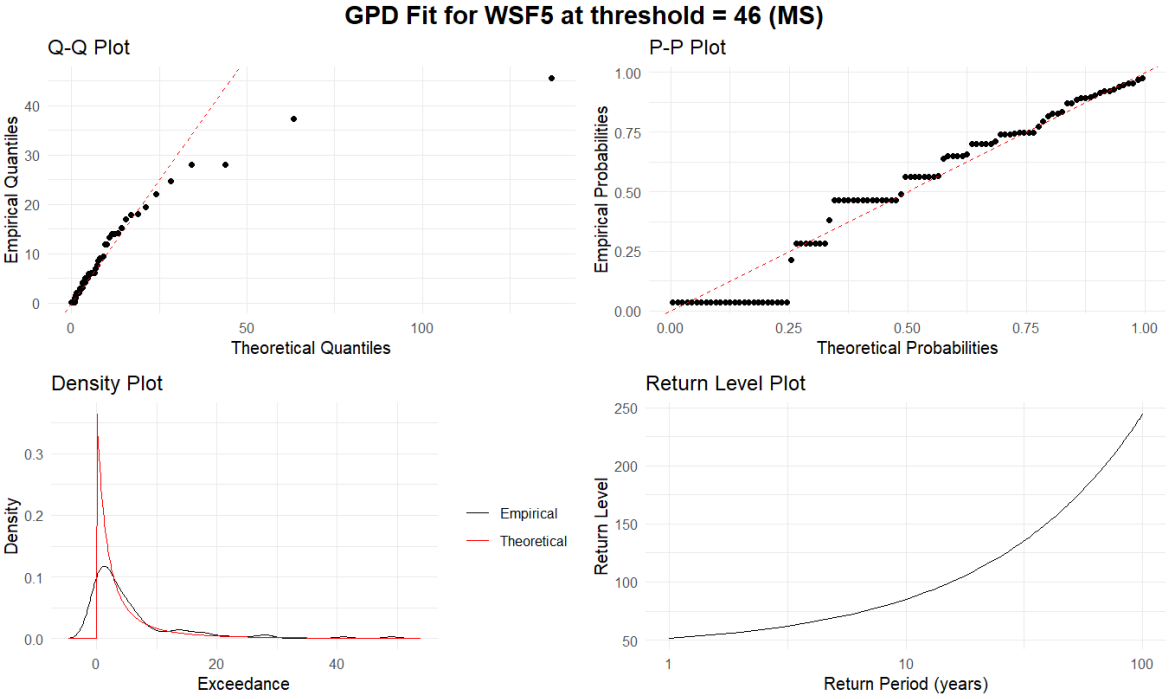

Figure 4.33 GPD Distribution Diagnostic Plots: Wind Speed in Mississippi

Figure 4.30, Figure 4.31 and *Figure 4.32* show that while the GPD model generally provides an acceptable fit to the data, there are significant concerns regarding its ability to accurately predict the most extreme wind speed events, although the model shows an underestimation of the severity of extreme events, particularly at higher quantiles and longer return periods. This underestimation is consistent across all three states but is notably pronounced in Texas, where the extreme values exhibit substantial deviation from the theoretical predictions.

Flood Level

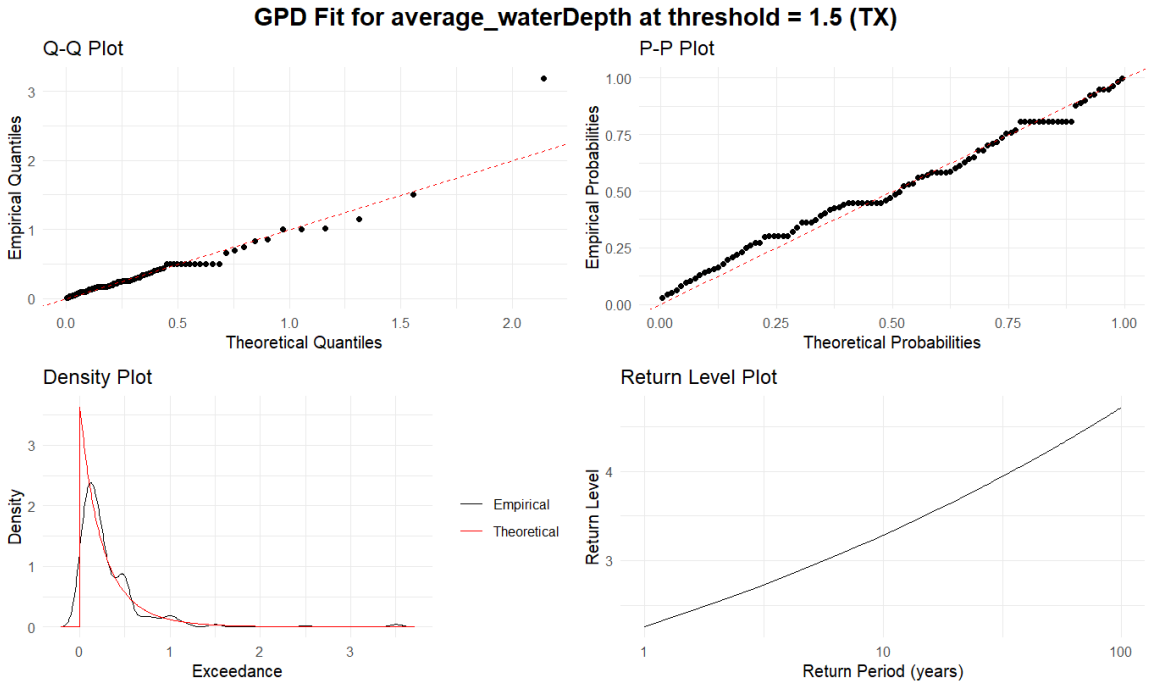

Figure 4.34 GPD Distribution Diagnostic Plots: Flood Level in Texas

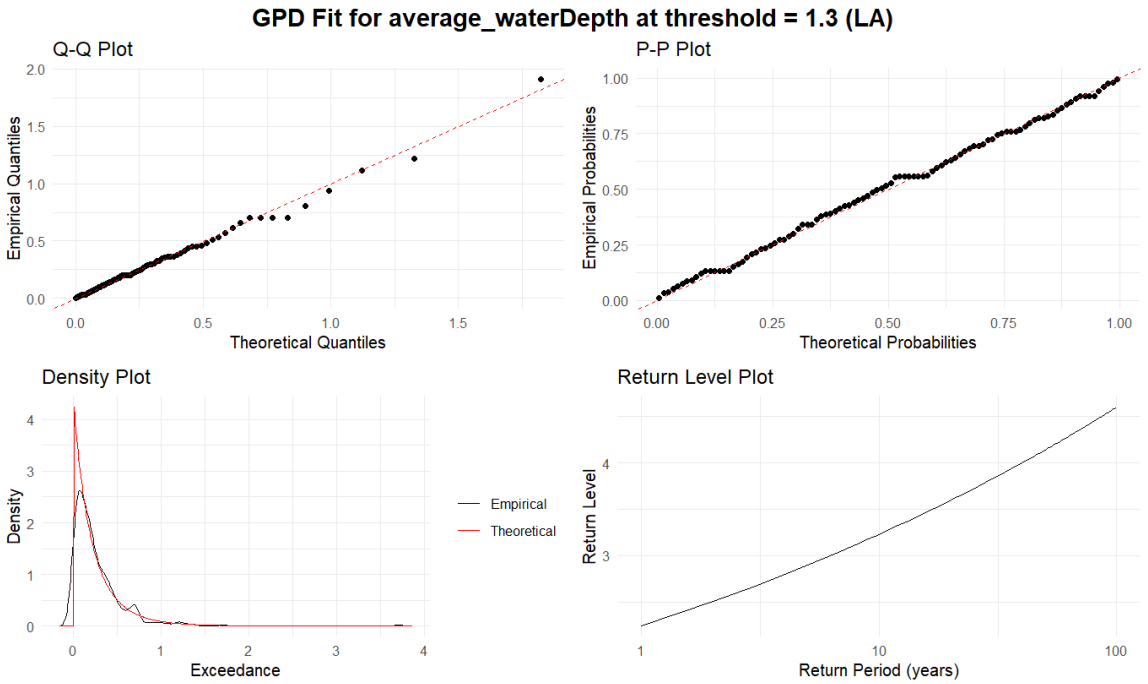

Figure 4.35: GPD Distribution Diagnostic Plots: Flood Level in Louisiana

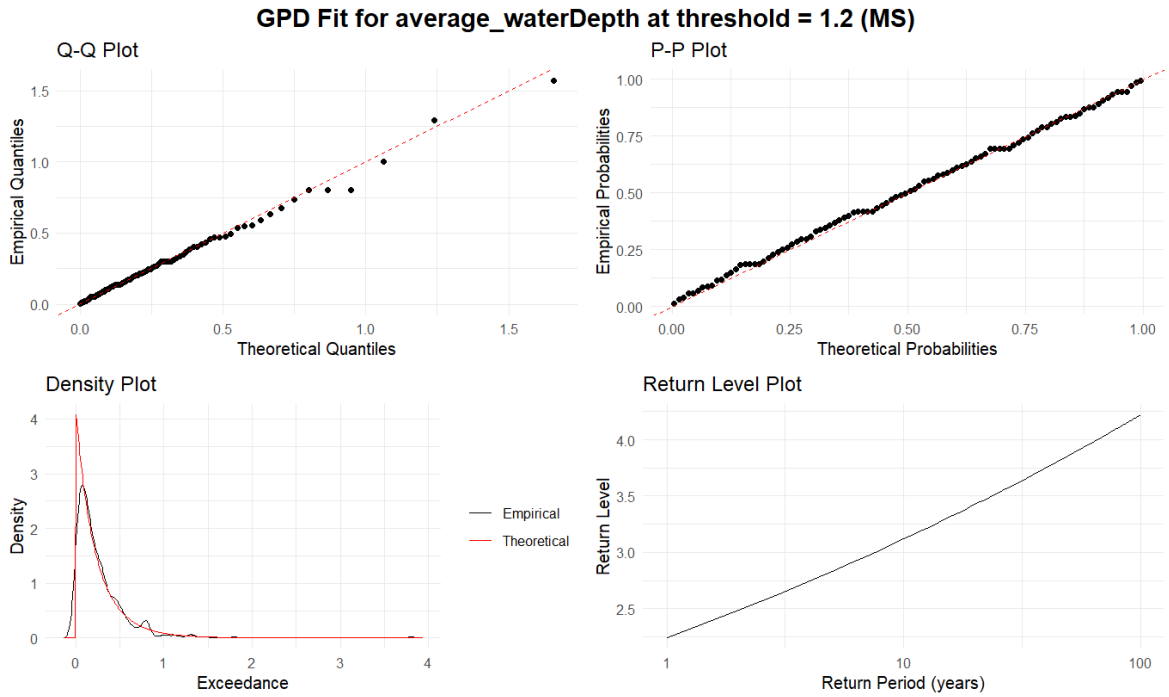

Figure 4.36: GPD Distribution Diagnostic Plots: Flood Level in Mississippi

Figure 4.34, Figure 4.35 and *Figure 4.36* indicate that the GPD model generally provides a good fit to the data, but there are some differences when it comes to predicting the most extreme flood events. Specifically, the model shows a tendency to underestimate the severity of extreme flood levels at higher quantiles and longer return periods. This underestimation is consistent across all three states but appears to be slightly more pronounced in Texas.

5.1.2.4 Return Level (GPD)

Final 50-year Return Levels:

	TX	LA	MS
Rainfall Level (in)	8.996515	9.889784	10.02712
Wind Speed (mph)	15682.18	89.72131	169.2415
Flood Level (in)	4.235453	4.139623	3.862646

Table 4.16: 50-Year Return Levels for Rainfall, Wind Speed, and Flood Levels (GPD)

The 50-year return levels in *Table 4.16* indicate that Mississippi is at the greatest risk for extreme rainfall and Texas is at the greatest risk for flooding, with the highest estimated values of 10.027 inches and 4.235 inches, respectively. Though slightly less vulnerable to rainfall, Louisiana shows significant risks for both wind and flood events. Texas and Mississippi, despite a potential outlier in wind speed, also face considerable risks, particularly in terms of flood levels.

5.1.2.5 Profile Log-likelihood Plots (GPD)

Rainfall

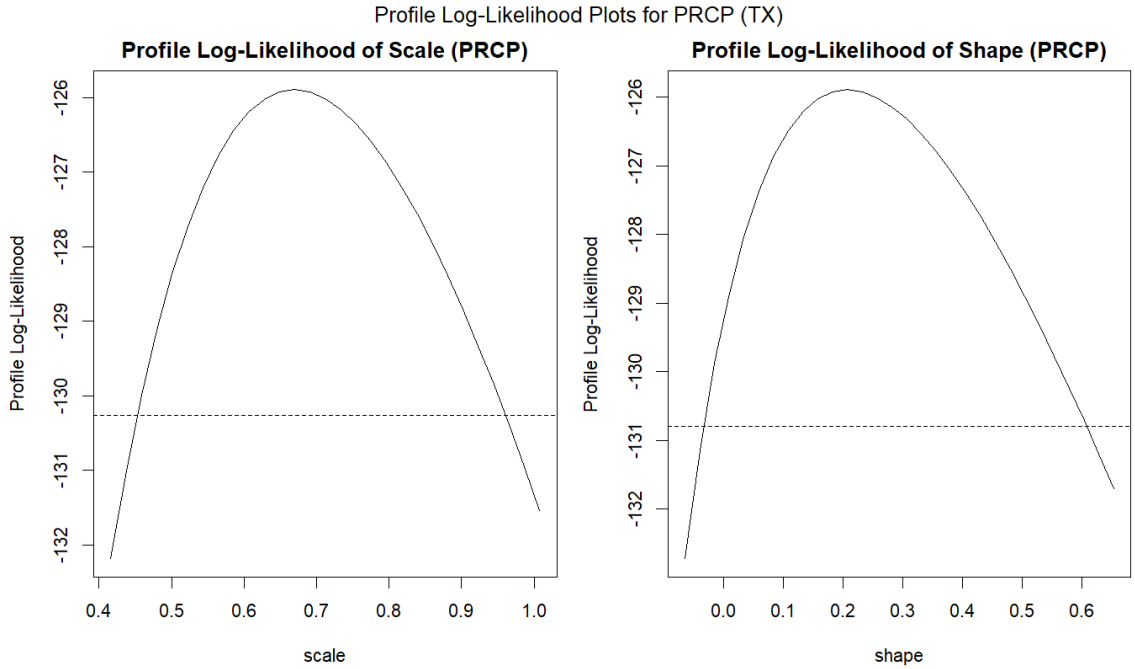

Figure 4.37: GPD Profile Log-Likelihood Plots: Rainfall Level in Texas

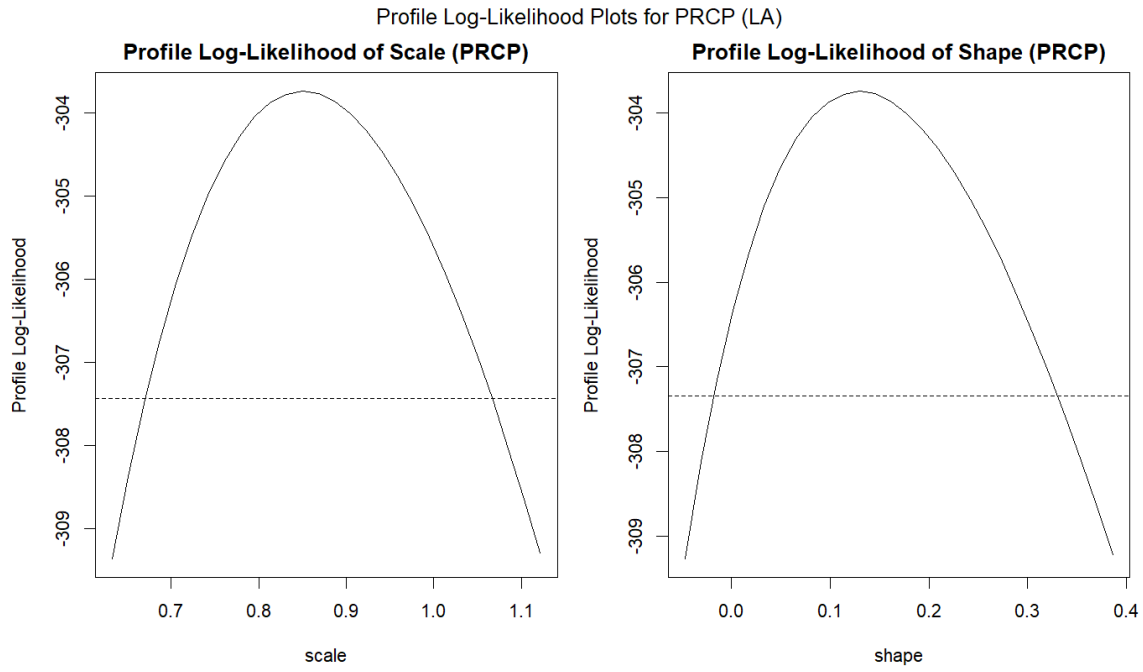

Figure 4.38: GPD Profile Log-Likelihood Plots: Rainfall Level in Louisiana

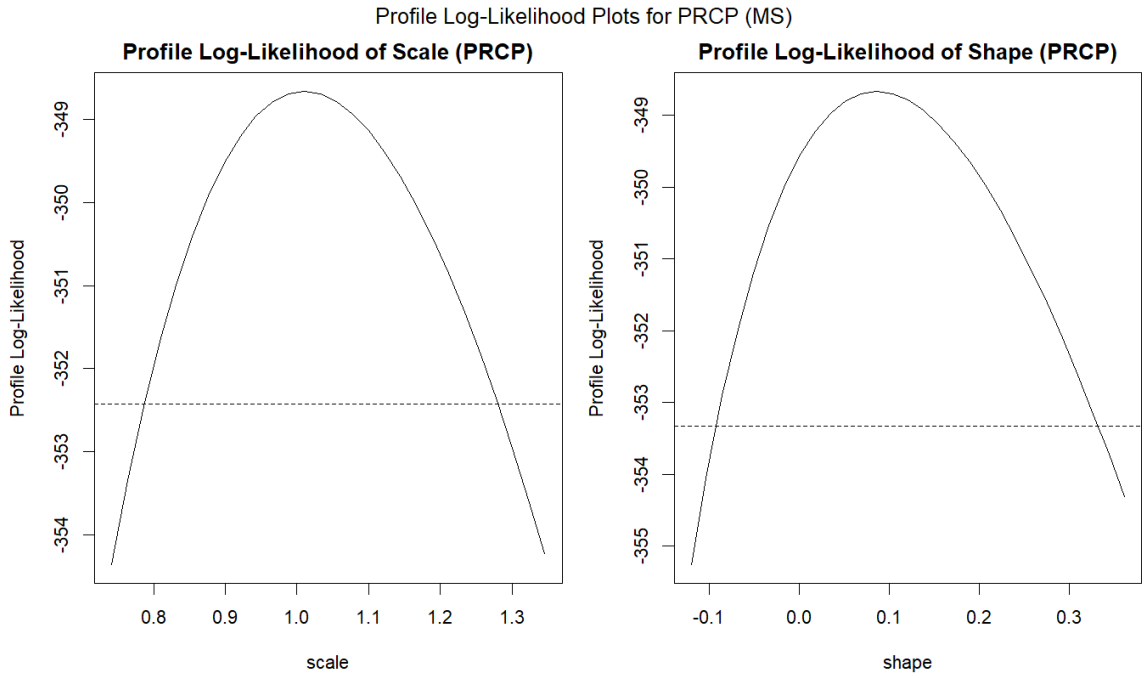

Figure 4.39: GPD Profile Log-Likelihood Plots: Rainfall Level in Mississippi

Confident Interval

	TX		LA		MS	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Scale	0.5042812	0.8354232	0.708411	0.9920075	0.8349	1.1855
Shape	0.0139591 1	0.4014292	0.003906	0.2552326	-0.04829	0.22096

Table 4.17: Confidence Intervals of Rainfall GPD Parameters

Wind Speed

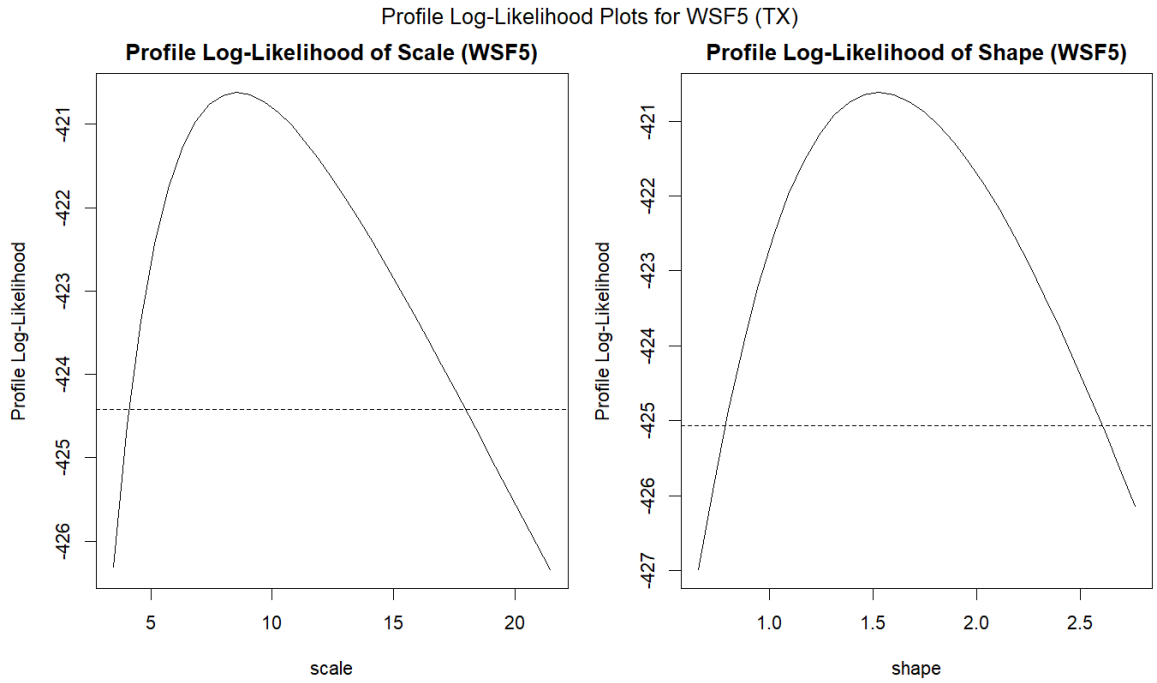

Figure 4.40: GPD Profile Log-Likelihood Plots: Windspeed in Texas

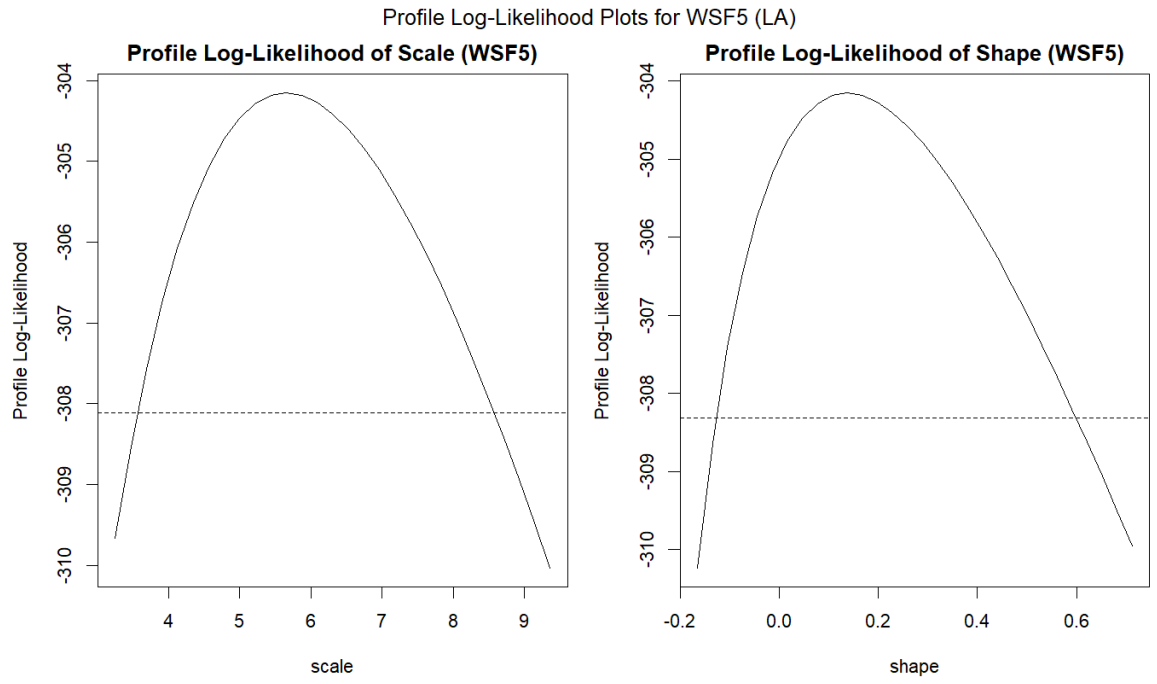

Figure 4.41: GPD Profile Log-Likelihood Plots: Windspeed in Louisiana

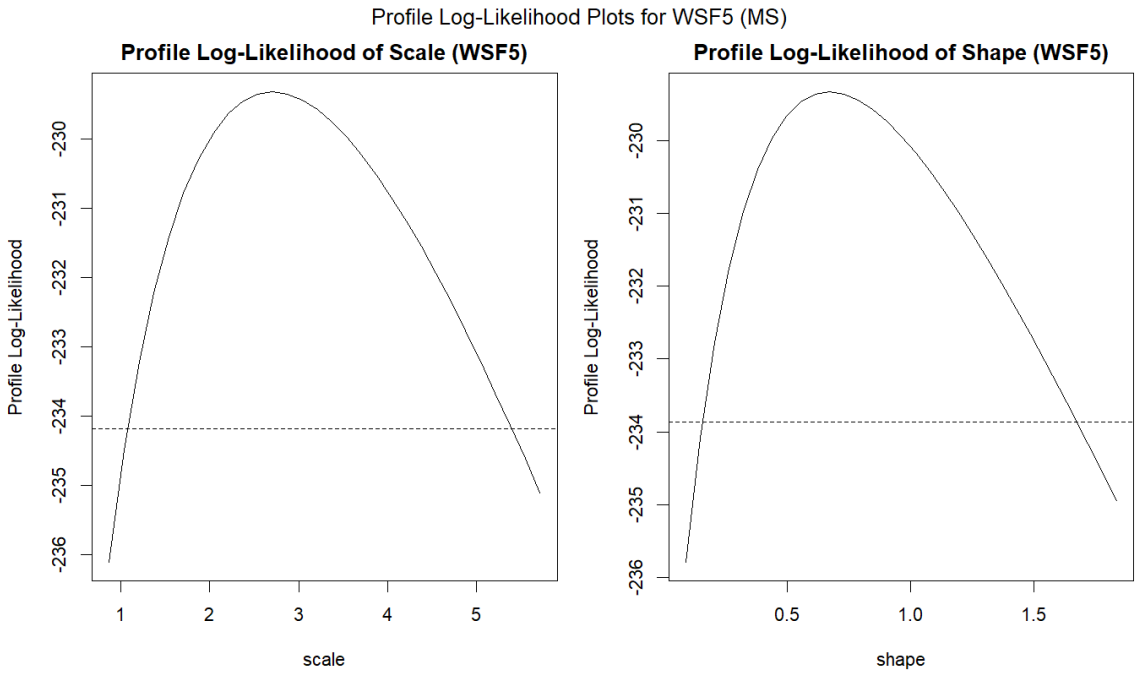

Figure 4.42: GPD Profile Log-Likelihood Plots: Windspeed in Mississippi

Confident Interval

	TX		LA		MS	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Scale	4.125571	12.937477	3.939481	7.362963	1.40050	4.0154
Shape	0.9609134	2.099315	-0.09964	0.374508	0.2137192	1.1270031

Table 4.18: Confidence Intervals of Windspeed GPD Parameters

Flood Level

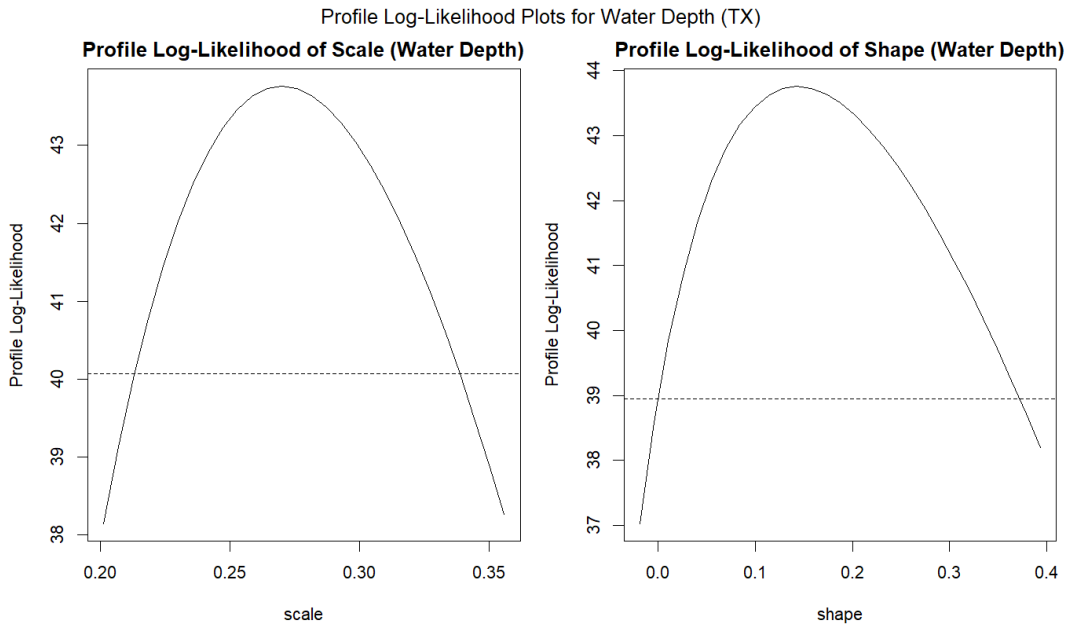

Figure 4.43: GPD Profile Log-Likelihood Plots: Flood Level in Texas

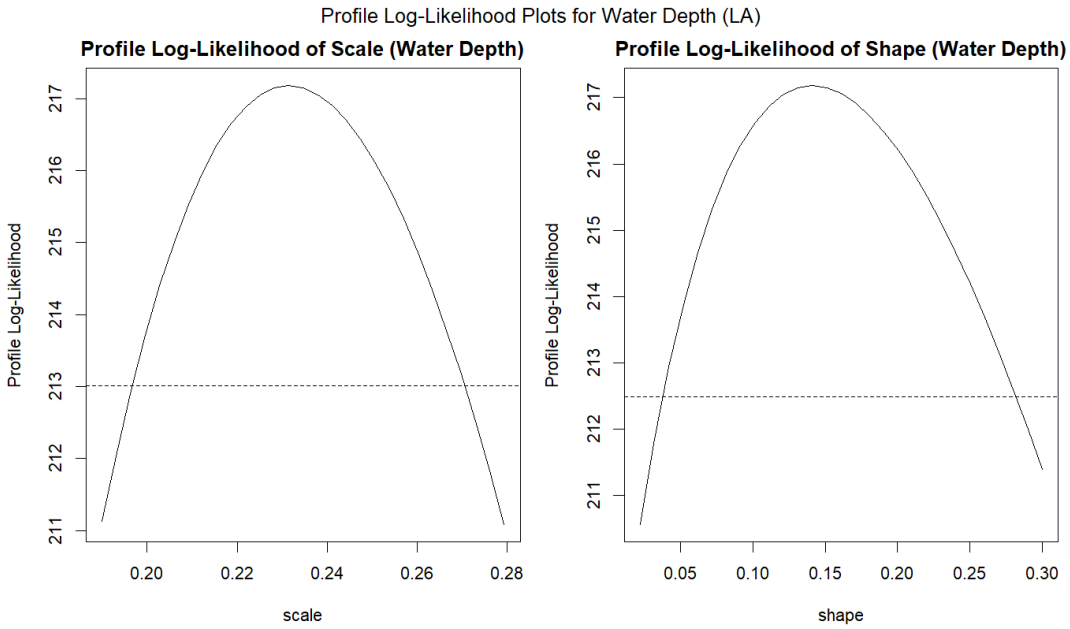

Figure 4.44: GPD Profile Log-Likelihood Plots: Flood Level in Louisiana

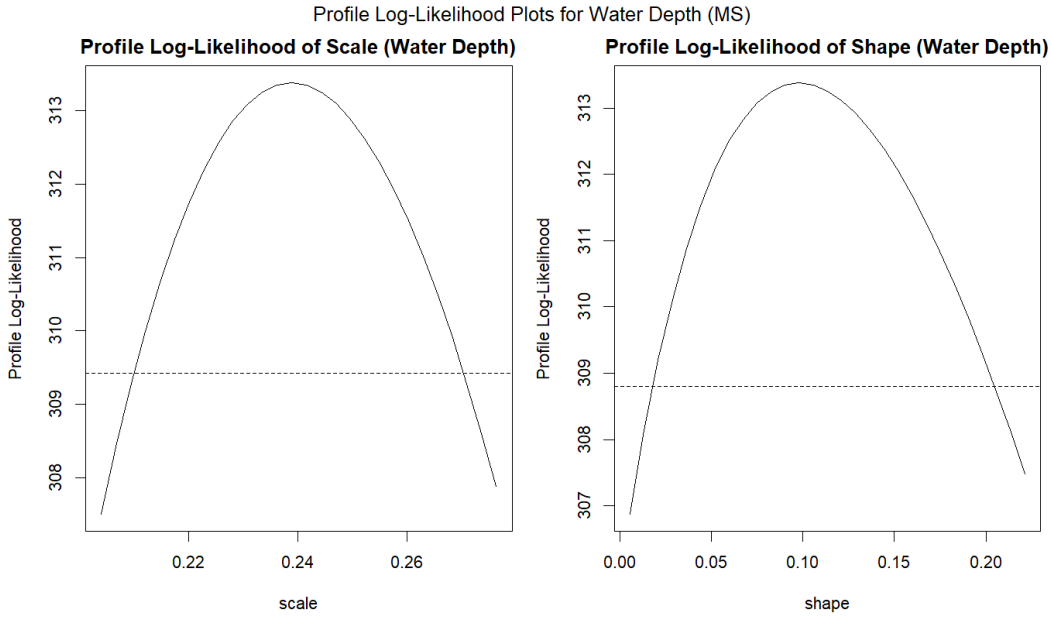

Figure 4.45: GPD Profile Log-Likelihood Plots: Flood Level in Mississippi

Confident Interval:

	TX		LA		MS	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Scale	0.2250979	0.3147815	0.2064852	0.2564390	0.21786	0.2598
Shape	0.02790215	0.25846326	0.06328365	0.21896584	0.037509	0.1584

Table 4.19: Confidence Intervals of Flood Level GPD Parameters

The profile log-likelihood plots for rainfall, windspeed, and flood levels across Texas (TX), Louisiana (LA), and Mississippi (MS) exhibit concave shapes, indicating that the likelihood functions for the scale and shape parameters possess single maxima. This concavity strongly indicates that the estimated parameters are stable, with the optimisation process having successfully converged to a unique solution. The peaks of these plots represent the most likely values of the parameters under the fitted Generalised Pareto Distribution (GPD) models.

The absence of steep slopes and sharp peaks in the profile log-likelihood plots (*Figure 4.37, Figure 4.38, Figure 4.39, Figure 4.40, Figure 4.41, Figure 4.42, Figure 4.43, Figure 4.44* and *Figure 4.45*) indicates that while the model appears to capture the overall structure of the data, it may not be capturing all the finer details with high accuracy. From *Tables 4.17, Table 4.18* and *Table 4.19*, narrower confidence intervals, particularly for location parameters, suggest greater confidence in those estimates, whereas wider intervals for scale and shape parameters indicate higher uncertainty. The shape parameters determine whether the model anticipates the occurrence of very extreme events (heavy tails) or if such extremes are less likely (bounded tails). When the lower bound (2.5%) and upper bound (97.5%) have different signs, further analysis of the shape parameters is necessary to assess the tail behaviours.

4.1.2.6 Goodness of Fit (GPD)

Rainfall Level

	TX	LA	MS
Kolmogorov-Smirnov test	0.9542	0.9428	0.7248
Cramer-von Mises test	0.9827	0.9596	0.6996

Table 4.20: Goodness of Fit Results for Rainfall Level GPD Model

Windspeed

	TX	LA	MS
Kolmogorov-Smirnov test	0.00028	0.3024	0.0004851
Cramer-von Mises test	0.01654	0.4729	0.06386

Table 4.21: Goodness of Fit Results for Windspeed GPD Model

Flood Level

	TX	LA	MS

Kolmogorov-Smirnov test	0.07293	0.1719	0.08383
Cramer-von Mises test	0.119	0.4065	0.2623

Table 4.22: Goodness of Fit Results for Flood Level GPD Model

From *Table 4.20, Table 4.21* and *Table 4.22*, the GPD models generally provide a good fit for modelling extreme rainfall across all states, but they are not that good in windspeed and flood level modelling. The fit is particularly strong for rainfall in Texas and Louisiana, while the fit for windspeed in Texas and Mississippi is less convincing, indicating potential limitations of the GPD model in capturing the full range of extreme windspeed events in these states. The moderate fit observed for flood levels across all states suggests that the GPD models are suitable, although they might perform weaker in capturing the tail behaviour of the distributions in these cases.

4.1.3 Copula Analysis

The NAs in the table represent a failure to fit the specific type of copula model.

4.1.3.1 Copula Analysis (Texas)

GEV Parameters fitting:

Copula	AIC	BIC	Log-Likelihood	Kendall Tau	Spearman Rho
Gumbel	2.000000	3.295837	-1.91e-07	1.490116e-08	-0.001602055
Clayton	NA	NA	NA	NA	NA
Frank	NA	NA	NA	NA	NA
t	7.836962	9.132799	-2.918481	-5.751845e-02	-0.226265417

Table 4.23: GEV Copula Model Fit and Dependence Measures for Texas

In *Table 5.23*, the Gumbel copula shows the lowest AIC and BIC values, suggesting a slightly better fit compared to the t-copula. However, the log-likelihood values are close to zero, and the Kendall Tau and Spearman Rho values are nearly negligible, indicating a weak dependence between the variables under the GEV model.

GPD Parameters fitting:

Copula	AIC	BIC	Log-Likelihood	Kendall Tau	Spearman Rho
Gumbel	2.0000002	4.499810	-8.320793e-08	1.490116e-0	-0.001602055
Clayton	1.592915	4.092725	2.035422e-01	3.051453e-02	0.045753552
Frank	1.9935377	4.493347	3.231134e-03	3.199846e-03	0.004799756
t	-0.278926	2.220883	1.139463	1.261230e-02	NA

Table 4.24: GPD Copula Model Fit and Dependence Measures for Texas

Similarly, in *Table 4.24*, the GPD model results indicate that the *t* copula provides the best fit, with the lowest AIC and BIC values and the highest log-likelihood value. The *t* copula also shows the most substantial dependence, with Kendall Tau, although it shows NA in the Spearman Rho column, suggesting potential issues in estimating this measure.

Best Fitting Copula in Texas

Copula Type	Estimate of Rho (ρ)	Standard Error of Rho	*p-Value*

t (GPD)	0.01981	0.078	0.9895

Table 4.25: Best Fitting Copula for Texas

In *Table 4.25*, the *t* copula is identified as the best fitting copula for modelling the dependence structure of extreme weather events in Texas. The estimated Rho (ρ), which measures the strength of dependence between the variables, is 0.01981, with a standard error of 0.078. Although this small value of Rho suggests a weak dependence between the variables, the p-value associated with the goodness-of-fit test is 0.9895, which is very high. This indicates that the t-copula with parameter $\rho=0.01981$ is a suitable model for the data.

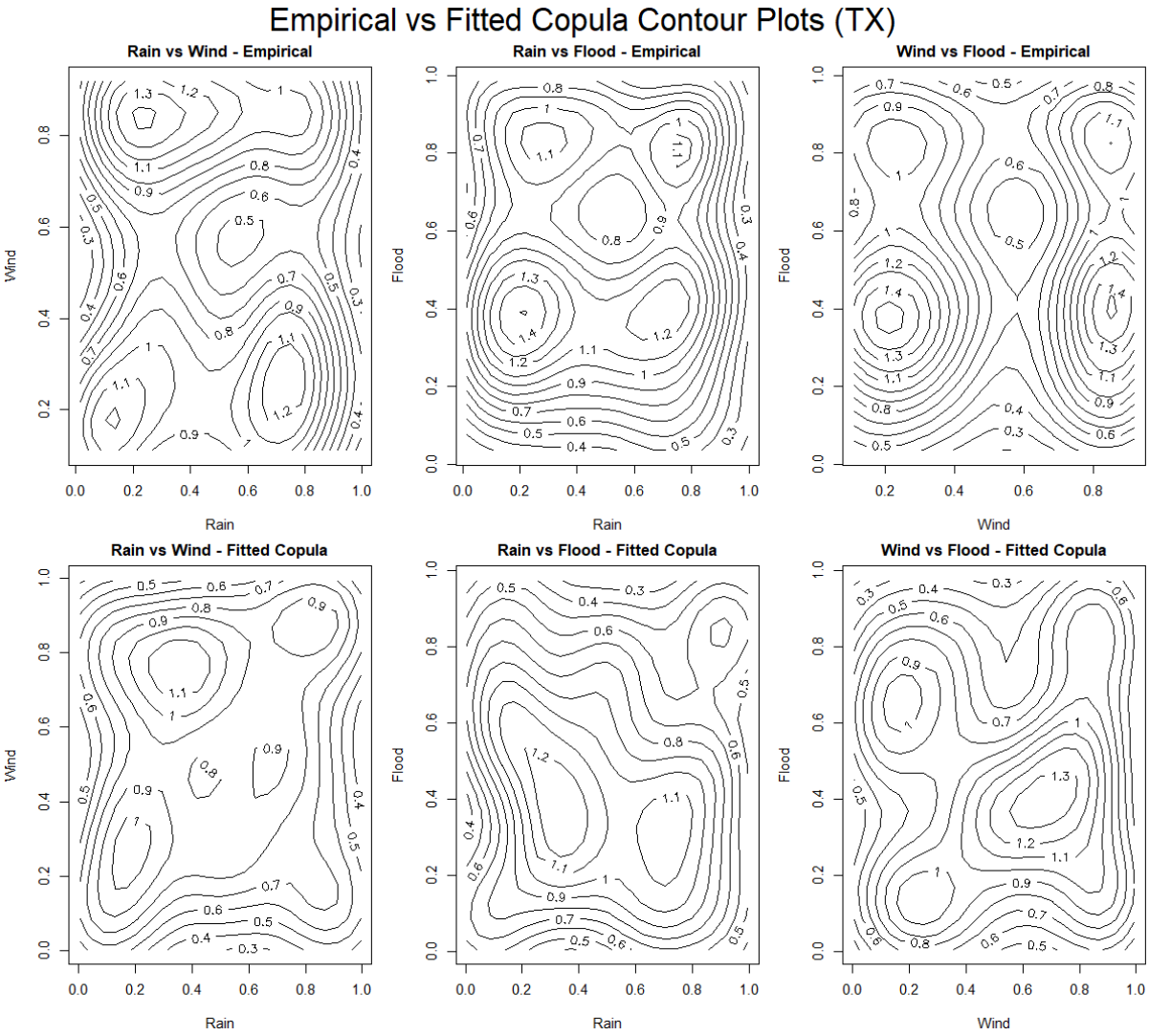

Figure 4.46: Empirical vs. Fitted Copula Contour Plots (Texas)

Figure 4.46 provides a comparison to visually assess how well the fitted copula model captures the dependence structure observed in the empirical data.

Rain vs Wind:

Empirical Plot: The contours suggest a complex dependence structure with multiple peaks and valleys, indicating higher and lower density areas where these variables co-occur.

Fitted Copula Plot: The contours are smoother and somewhat align with the empirical plot's general shape, but there are differences in the fine structure, indicating that while the copula model captures the broad dependence, it may miss some of the more complex interactions seen in the empirical data.

Rain vs Flood:

Empirical Plot: The plot indicates several regions of high density, particularly in the mid-range values, suggesting a moderate positive dependence between rainfall and flood levels.

Fitted Copula Plot: The fitted plot captures the general trend of the empirical data but with smoother transitions and less pronounced peaks. This suggests that the copula model approximates the overall relationship but may not fully capture the more nuanced patterns in the data.

Wind vs Flood:

Empirical Plot: The contours are spread out, indicating a weaker and possibly more dispersed dependence structure between these variables.

Fitted Copula Plot: The fitted model again smooths out the contours and captures the general shape of the empirical distribution. However, the copula model may miss some finer details and variability in the empirical data.

The comparison of empirical and fitted copula contour plots in Texas reveals that the fitted copula models generally capture the broad trends and dependence structures seen in the empirical data. However, the models tend to smooth out the contours, potentially overlooking some more complex or subtle interactions in the empirical joint distributions.

This smoothing effect is particularly evident in the Rain vs Wind and Rain vs Flood comparisons, where the fitted copula models approximate the overall structure but lack the detailed peaks and troughs observed empirically.

4.1.3.1 Copula Analysis (Louisiana)

GEV Parameters fitting:

Copula	AIC	BIC	Log-Likelihood	Kendall Tau	Spearman Rho
Gumbel	2.000000	3.332205	-3.980e-09	1.4901e-08	-0.00160
Clayton	2.000000	3.332205	0	0	-0.0000258
Frank	1.962216	3.294420	1.8892e-02	1.1529e-02	0.017293596
t	2.955009	4.287213	-4.775e-01	-1.9236e-02	0.346968367

Table 4.26: GEV Copula Model Fit and Dependence Measures for Louisiana

In *Table 4.26*, the Frank copula shows the lowest AIC (1.962) and BIC (3.294) values, indicating a slightly better fit compared to the other copulas under the GEV model for Louisiana. The log-likelihood value, although positive, is close to zero, reflecting a minimal improvement in fit. However, the Kendall Tau and Spearman Rho values are also very low, suggesting weak dependence between the variables. The Gumbel and Clayton copulas have identical AIC and BIC values (2.000 and 3.332, respectively) but show negligible dependence, as indicated by their near-zero Kendall Tau and Spearman Rho values. The *t* copula exhibits the highest AIC (2.955) and BIC (4.287) values, with a negative log-likelihood, indicating the poorest fit among the copulas, despite showing a higher Spearman Rho value, which may suggest some level of dependence but with a more complex structure.

GPD Parameters fitting:

Copula	AIC	BIC	Log-Likelihood	Kendall Tau	Spearman Rho
Gumbel	2.000002	4.663441	-8.094816e-07	1.490116e-08	-0.001602055
Clayton	NA	NA	NA	NA	NA
Frank	NA	NA	NA	NA	NA
t	22.396333	25.059772	-1.0198e+01	-2.229628e-02	NA

Table 4.27: GPD Copula Model Fit and Dependence Measures for Louisiana

Similarly, in *Table 4.27*, the GPD model results show that the Gumbel copula provides a better fit among the few copulas tested, with an AIC of 2.000 and a BIC of 4.663. The log-likelihood is again nearly zero, indicating a marginal fit. The Kendall Tau is negligible, and Spearman Rho is negative, indicating very weak or potentially no meaningful dependence. The t-copula, with much higher AIC (22.396) and BIC (25.060) values and a significantly negative log-likelihood, suggests a poor fit for the GPD model. The Kendall Tau is slightly negative, and the Spearman Rho is not available (NA), pointing to potential issues in estimating this measure, similar to what was seen in the GEV model.

Best Fitting Copula in Louisiana

Copula Type	Estimate of alpha (α)	Standard Error of alpha	*p-Value*
Gumbel (GPD)	1	0.043	0.7847

Table 4.28: Best Fitting Copula for Louisiana

From *Table 4.28*, the estimate of alpha being 1, combined with a high *p-value* and a low standard error, indicates that the dependence between the variables is likely very weak or

negligible. This suggests that the Gumbel copula may not be capturing a meaningful dependence structure for the data in Louisiana. As a result, the Gumbel copula, while the best-fitting copula for this dataset, may not provide substantial insights into the relationships between the variables due to the apparent lack of significant dependence.

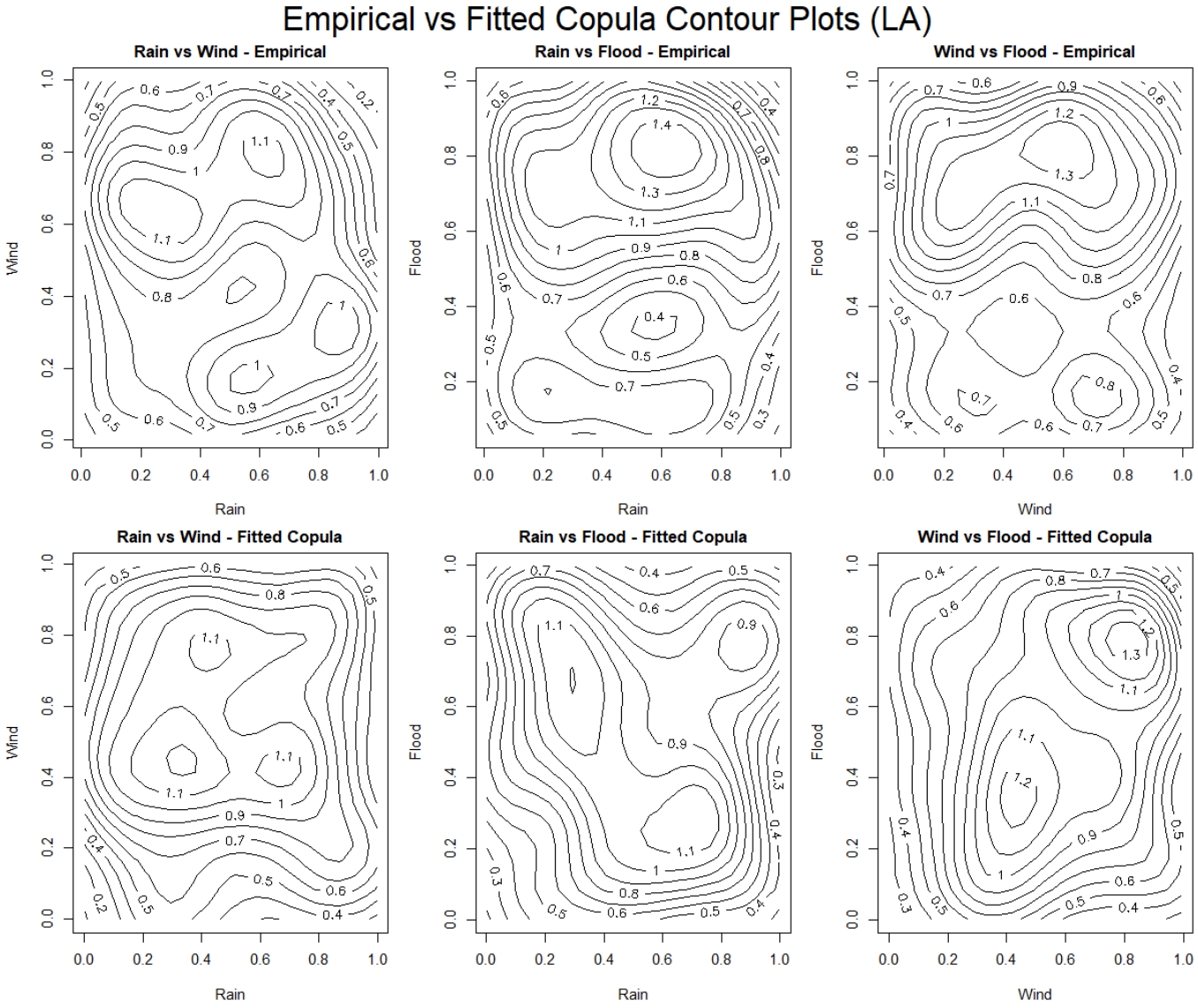

Figure 4.47: Empirical vs. Fitted Copula Contour Plots (Louisiana)

Figure 4.47 This figure compares the empirical joint distributions (top row) and the fitted copula model distributions (bottom row) for three variables: Rain vs Wind, Rain vs Flood, and Wind vs Flood in Louisiana.

Rain vs Wind:

Empirical Plot: The contours display a fairly complex structure with multiple peaks and valleys, indicating areas where rain and wind tend to co-occur with varying intensities.

Fitted Copula Plot: The fitted plot captures the general pattern of the empirical data, including the major peaks, but with smoother transitions and less pronounced variability. This suggests that while the fitted copula model approximates the overall dependence, it may miss some finer details in the empirical data.

Rain vs Flood:

Empirical Plot: The contours indicate multiple regions of high density, particularly in the mid to higher values, suggesting a moderate positive dependence between rainfall and flood levels.

Fitted Copula Plot: The fitted plot closely follows the structure of the empirical contours, capturing the major trends and peaks, though with some smoothing. This indicates that the fitted copula model effectively captures the overall relationship between rain and flood, but again, with less detail than the empirical data might suggest.

Wind vs Flood:

Empirical Plot: The contours are spread out, suggesting a weaker and possibly more diffuse dependence between these variables.

Fitted Copula Plot: The fitted model successfully captures the broad patterns seen in the empirical plot, including the general direction and shape of the contours. However, as with the other pairs, the fitted copula smooths out the details, potentially overlooking some of the complexity in the empirical data.

The comparison of empirical and fitted copula contour plots in Louisiana reveals that the fitted copula models generally capture the main trends and dependence structures seen

in the empirical data. The fitted plots replicate the overall shapes and directions of the empirical contours, indicating that the models are reasonably good at approximating the joint distributions of the variables. However, the fitted copulas tend to smooth out the finer details, which may result in a less accurate representation of the more complex interactions between variables. This smoothing effect is consistent across all three variable pairs. It suggests that while the fitted copulas provide a useful approximation, they might not fully capture all nuances of the empirical relationships.

4.1.3.1 Copula Analysis (Mississippi)

GEV Parameters fitting:

Copula	AIC	BIC	Log-Likelihood	Kendall Tau	Spearman Rho
Gumbel	1.9175958	3.175692	0.0412021	0.02678382	0.03802938
Clayton	0.8072453	2.065342	0.5963774	0.07193306	0.10762302
Frank	1.3567732	2.614870	0.3216134	0.05938728	0.08899598
t	3.2533759	4.511472	-0.626688	0.05973812	0.24678745

Table 4.29: GEV Copula Model Fit and Dependence Measures for Mississippi

In *Table 4.29*, the Clayton copula shows the lowest AIC (0.807) and BIC (2.065) values, suggesting it provides the best fit among the tested copulas under the GEV model for Mississippi. The log-likelihood for the Clayton copula is also the highest (0.596), further supporting its superior fit. Additionally, the Clayton copula displays the highest Kendall Tau (0.0719) and Spearman Rho (0.1076) values, indicating a modest level of dependence between the variables.

GPD Parameters fitting:

Copula	AIC	BIC	Log-Likelihood	Kendall Tau	Spearman Rho
Gumbel	2.000001	4.454349	-7.185e-07	1.490116e-08	-0.001602055
Clayton	NA	NA	NA	NA	NA
Frank	NA	NA	NA	NA	NA
T	1.477225	3.931572	2.613877e-01	-6.068932e-02	NA

Table 4.30: GPD Copula Model Fit and Dependence Measures for Mississippi

In *Table 4.30*, the *t* copula provides the best fit for the GPD model in Mississippi, with an AIC of 1.477 and a BIC of 3.932. The log-likelihood (0.2614) is positive, indicating a reasonable fit. However, the Kendall Tau value (-0.0607) is negative, which could indicate an inverse or complex relationship between the variables. The Spearman Rho value is not available (NA), suggesting potential difficulties in estimating this measure, as seen in previous analyses.

Best Fitting Copula in Mississippi

Copula Type	Estimate of Rho (ρ)	Standard Error of Rho	p-Value
t (GPD)	-0.09519	0.061	0.8307

Table 4.31: Best Fitting Copula for Mississippi

From the numbers in *Table 4.31*, the *t* copula is the best fitting copula for the data in Mississippi, but the estimated Rho value suggests a weak and potentially insignificant inverse relationship between the variables. The high *p-value* and small standard error imply that while the estimate is precise, the dependence it represents is not statistically significant.

This suggests that the dependence structure between these variables might belong to a more complex structure, and the variables may be largely independent in practical terms.

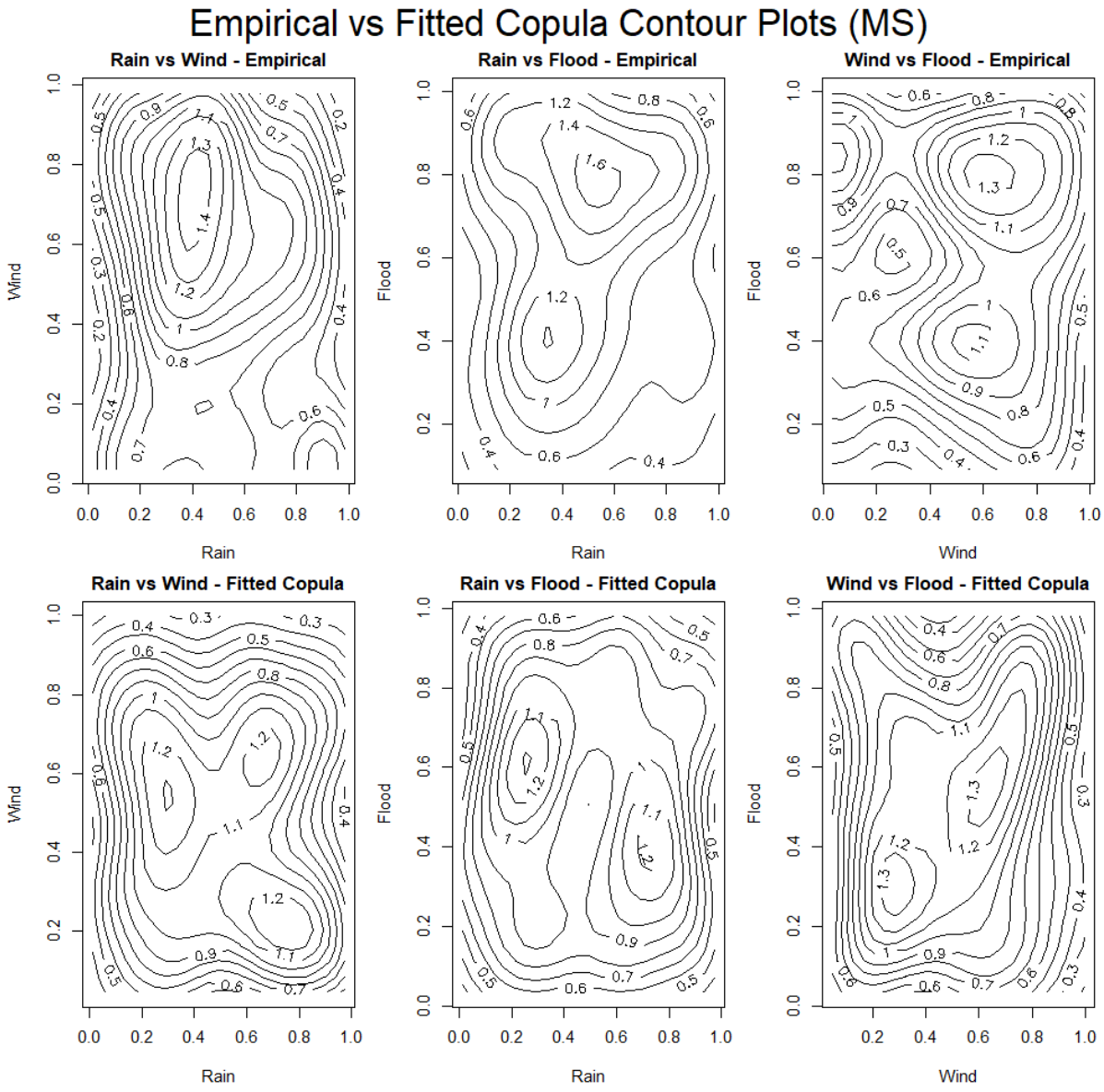

Figure 4.48: Empirical vs. Fitted Copula Contour Plots (Mississippi)

Figure 4.48 presents a comparison between the empirical joint distributions (top row) and the fitted copula model distributions (bottom row) for three variable pairs: Rain vs Wind, Rain vs Flood, and Wind vs Flood in Mississippi.

Rain vs Wind:

Empirical Plot: The contours are tightly packed and show a complex, strong dependence structure with distinct areas of high density. This suggests a significant relationship between rain and wind in Mississippi.

Fitted Copula Plot: The fitted contours generally follow the empirical pattern, but they are smoother and less complex. While the fitted copula captures the broad trends, it may oversimplify some of the intricacies observed in the empirical data.

Rain vs Flood:

Empirical Plot: The contours show regions of high density, especially at moderate to high levels of rain and flood, suggesting a strong positive dependence between these variables.

Fitted Copula Plot: The contours align reasonably well with the empirical data, capturing the main density regions. However, the fitted model smooths out some detailed variations, indicating that while the copula model captures the general dependence, it may not fully reflect all the subtleties of the empirical relationship.

Wind vs Flood:

Empirical Plot: The contours suggest a more diffuse, weaker dependence structure, with areas of moderate density indicating some relationship between wind and flood levels.

Fitted Copula Plot: The fitted contours capture the general shape and direction of the empirical contours but with smoother transitions and less distinct peaks. This suggests that the fitted copula model can approximate the overall dependence but may miss some finer details in the empirical data.

The comparison between the empirical and fitted copula contour plots for Mississippi reveals that the fitted copula models generally capture the main trends and dependence structures observed in the empirical data. However, the fitted copulas tend to smooth out the contours, potentially missing some of the more complex or subtle interactions in the empirical data. This effect is particularly noticeable in the Rain vs Wind and Rain vs Flood comparisons, where the fitted copula models approximate the overall patterns but with less detail. Overall, the fitted copulas provide a reasonable approximation of the joint

distributions, though they may not fully capture all the nuances of the empirical relationships in Mississippi.

4.1.4 Prediction

Texas

Windspeed = 50 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.336892	2.086110	0.9141347	1.626033	4.505272
Flood Level	1.785189	1.686074	0.3176210	1.509635	2.611602

Table 4.32: Statistics for Rainfall and Flood Levels at 50 mph Windspeed in Texas

Windspeed = 100 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.385260	2.091460	0.9102006	1.615186	4.822277
Flood Level	1.800821	1.697857	0.3467885	1.510655	2.606777

Table 4.33: Statistics for Rainfall and Flood Levels at 100 mph Windspeed in Texas

Windspeed = 150 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.421805	2.095086	0.9819626	1.622791	5.140716
Flood Level	1.829890	1.700297	0.3716445	1.506994	2.847258

Table 4.34: Statistics for Rainfall and Flood Levels at 150 mph Windspeed in Texas

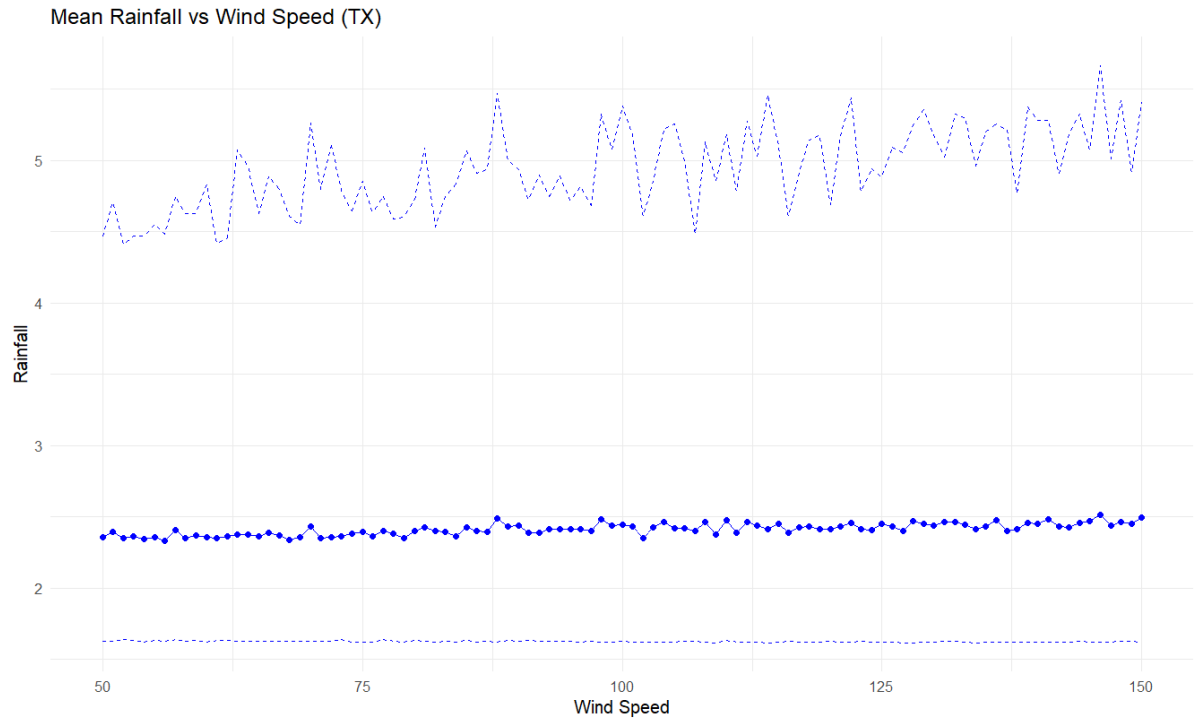

Figure 4.49: Mean Rainfall vs. Wind Speed in Texas

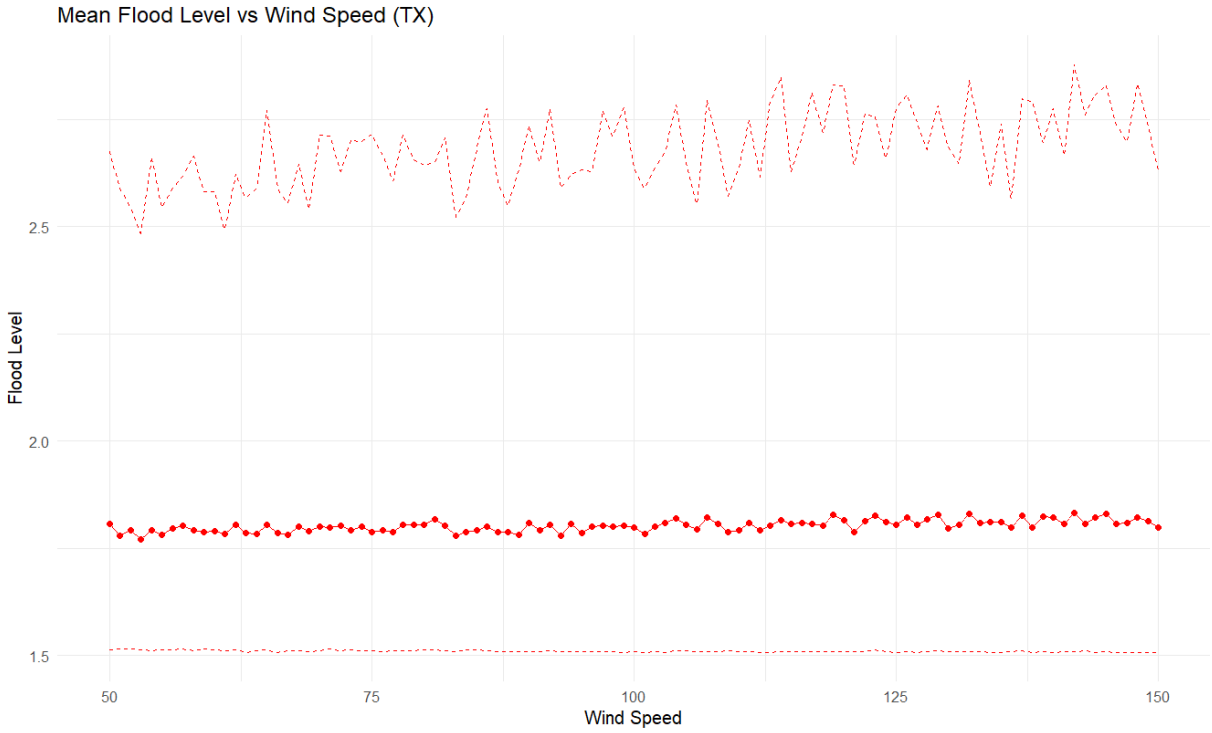

Figure 4.50: Mean Flood Level vs. Wind Speed in Texas

Table 4.32, Table 4.33 and *Table 4.34* present the estimated rainfall level and flood level in Texas under windspeed 50 mph, 100 mph and 150 mph, respectively. *Figure 4.49* and *Figure 4.50* present the relationship between wind speed and both rainfall and flood levels in Texas, with higher wind speeds associated with slightly greater mean values and increased variability in rainfall and flood levels.

Louisiana

Windspeed = 50 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.354263	1.981946	1.3582859	1.533312	4.305792
Flood Level	1.553796	1.478658	0.2564752	1.309148	2.253055

Table 4.35: Statistics for Rainfall and Flood Levels at 50 mph Windspeed in Louisiana

Windspeed = 100 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.460493	2.150972	0.9672123	1.524437	5.443626
Flood Level	1.479474	1.412385	0.1819847	1.310474	1.949256

Table 4.36: Statistics for Rainfall and Flood Levels at 100 mph Windspeed in Louisiana

Windspeed = 150 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.250214	2.160442	0.6506854	1.525330	3.961199
Flood Level	1.646784	1.548682	0.3081647	1.339864	2.135181

Table 4.37: Statistics for Rainfall and Flood Levels at 150 mph Windspeed in Louisiana

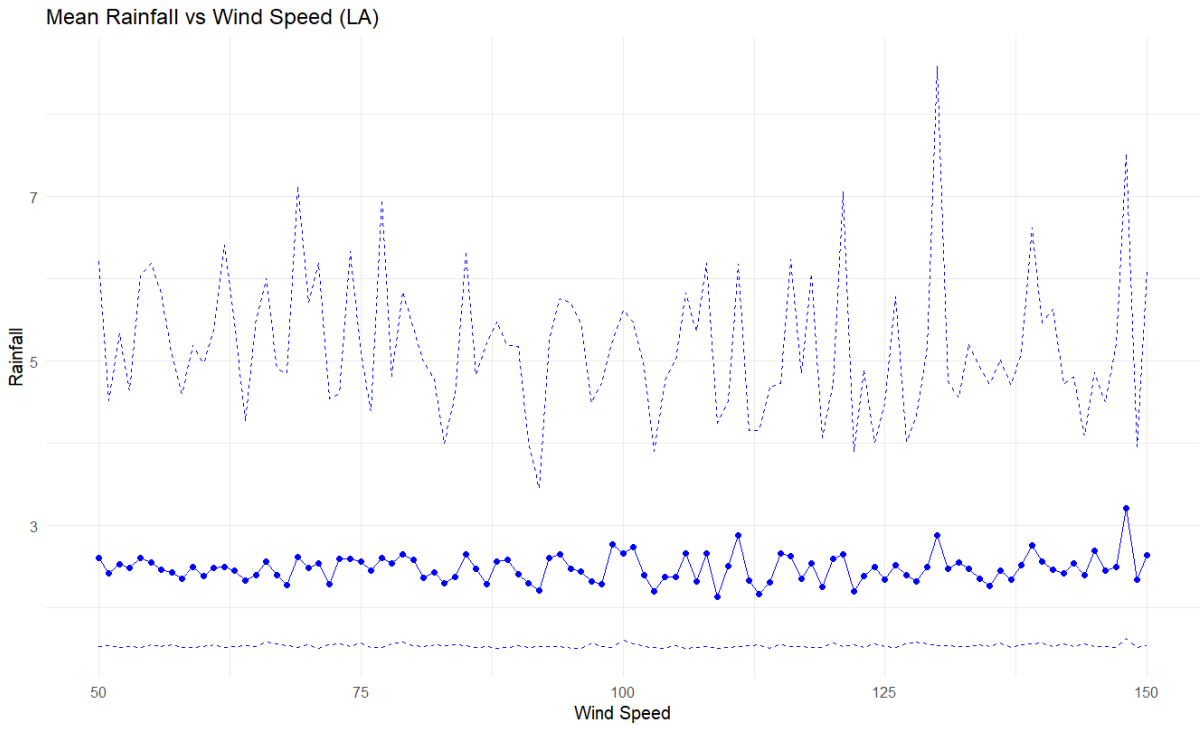

Figure 4.51: Mean Flood Level vs. Wind Speed in Louisiana

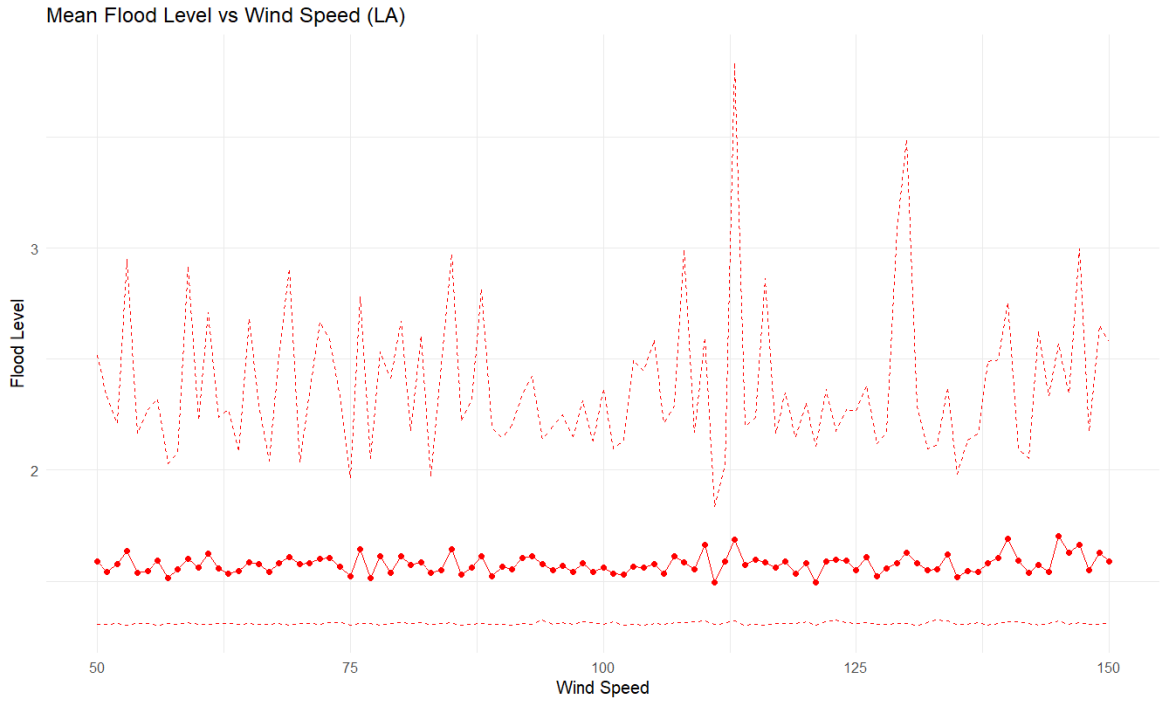

Figure 4.52: Mean Flood Level vs. Wind Speed in Louisiana

Table 4.35, Table 4.36 and *Table 4.37* display the estimated rainfall and flood levels in Louisiana at wind speeds of 50 mph, 100 mph, and 150 mph, respectively. *Figure 4.51* and *Figure 4.52* show the relationship between wind speed and both rainfall and flood levels in Louisiana, showing that as wind speeds increase from 50 mph to 150 mph, there is a slightly increase trend in mean values and greater variability in both rainfall and flood levels experience.

Mississippi

Windspeed = 50 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.275670	2.002273	0.9737421	1.334465	4.907456
Flood Level	1.436571	1.363909	0.2375550	1.205374	2.089311

Table 4.38: Statistics for Rainfall and Flood Levels at 50 mph Windspeed in Mississippi

Windspeed = 100 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.534261	1.822176	1.6738161	1.304676	7.149946
Flood Level	1.501591	1.324046	0.4405419	1.201695	2.771513

Table 4.39: Statistics for Rainfall and Flood Levels at 100 mph Windspeed in Mississippi

Windspeed = 150 mph

Variable	Mean	Median	SD	Lower CI	Upper CI
Rainfall	2.645477	1.715083	2.0388688	1.302468	8.250842
Flood Level	1.518626	1.305266	0.4616781	1.200609	2.878321

Table 4.40: Statistics for Rainfall and Flood Levels at 150 mph Windspeed in Mississippi

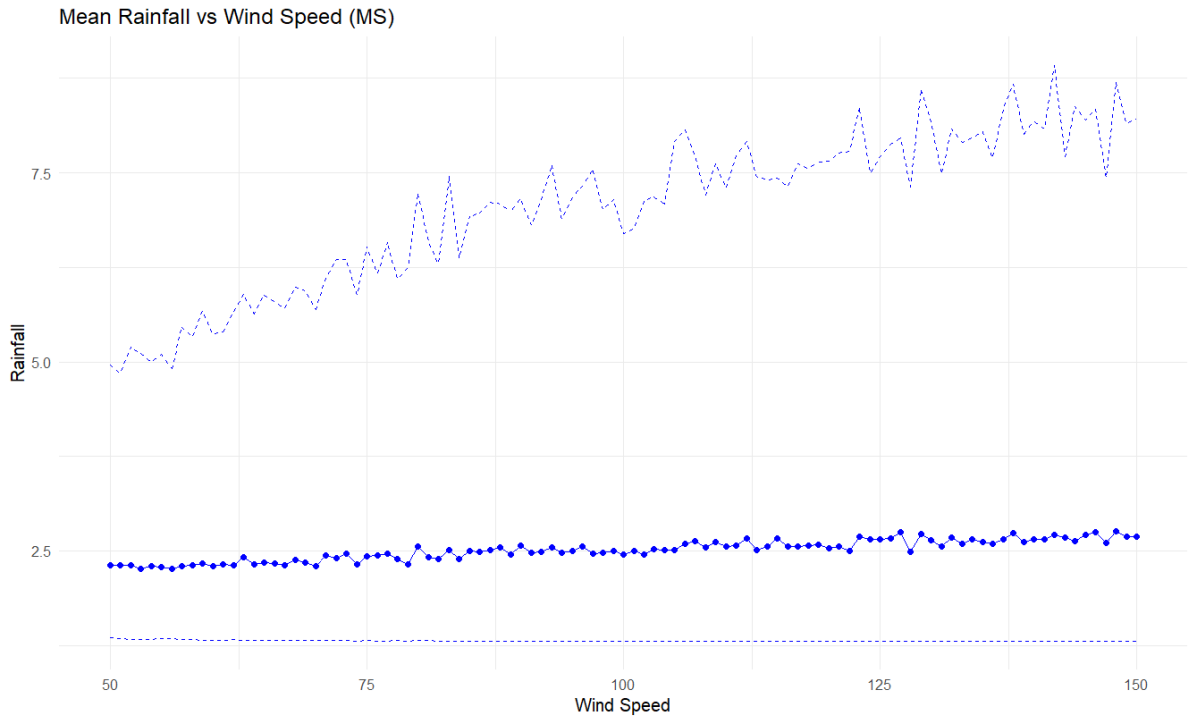

Figure 4.53: Mean Flood Level vs. Wind Speed in Mississippi

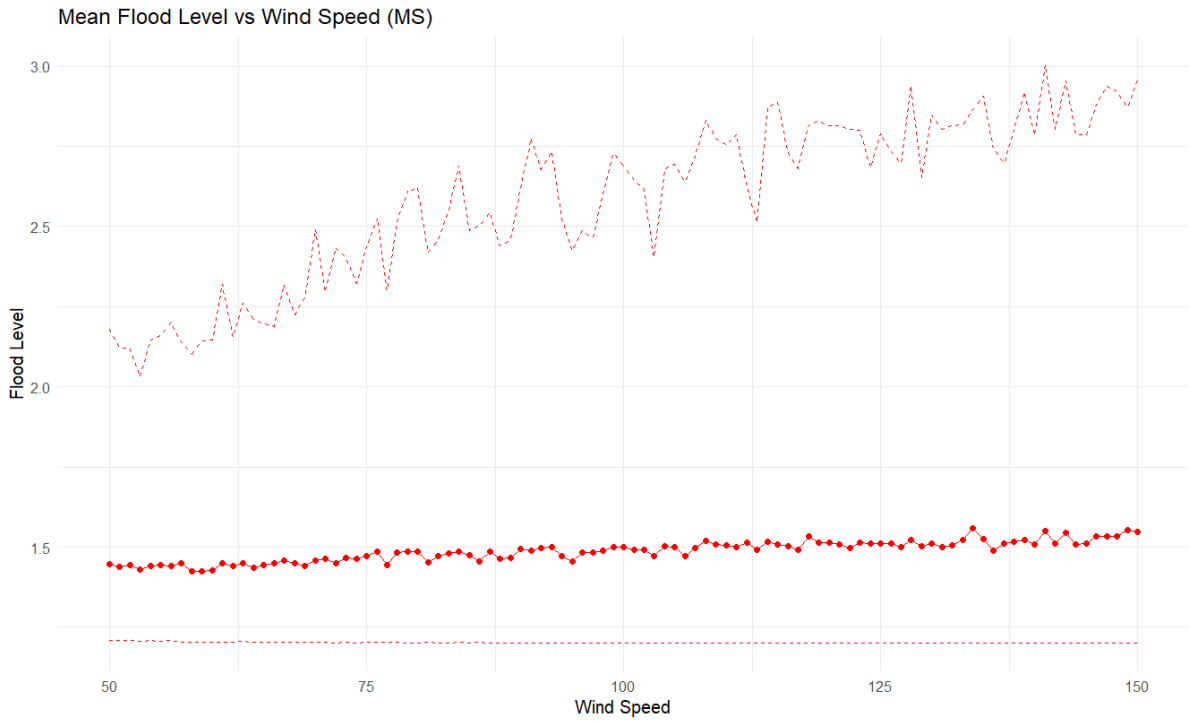

Figure 4.54: Mean Flood Level vs. Wind Speed in Mississippi

Table 4.38, Table 4.39 and *Table 4.40* show the estimated rainfall and flood levels in Mississippi at wind speeds of 50 mph, 100 mph, and 150 mph, respectively. *Figure 4.53* and *Figure 4.54* illustrate the relationship between wind speed and both rainfall and flood levels in Mississippi, demonstrating that as wind speeds increase from 50 mph to 150 mph, there is a slight rise in mean values along with greater variability in both rainfall and flood levels.

4.1.5 Annual Joint Probability

Rainfall Level (in)	Windspeed (mph)	Flood Level (in)
3	80	1.8

Table 4.41: Selected Threshold Scenario for Rainfall, Windspeed, and Flood Level

Table 4.41 presents the selected threshold scenario for calculating the annual joint exceedance probability of extreme weather events. Under this scenario, the probabilities of exceeding these thresholds for Texas, Louisiana and Mississippi are in *Table 5.42*.

Texas	Louisiana	Mississippi
0.0220	8e-04	0.0017

Table 4.42: Annual Joint Exceeding Probability

4.2. Vulnerability Model Results

4.2.1. Forward Selection Results

Texas (13,261 observations)

Selected variables:

elevatedBuildingIndicator, WSF5, numberOfFloorsInTheInsuredBuilding, PRCP, buildingAgeAtLoss, basementEnclosureCrawlspaceType, waterDepth

Fitted model

	Estimate	Pr(>|t|)
(Intercept)	10.2259916	< 2e-16
elevatedBuildingIndicator1	-0.4352655	< 2e-16
numberOfFloorsInTheInsuredBuilding	-0.0523679	2.28e-05
buildingAgeAtLoss	0.0025395	1.00e-10
basementEnclosureCrawlspaceType1	-0.3117258	< 2e-16
basementEnclosureCrawlspaceType2	-0.4244367	< 2e-16
basementEnclosureCrawlspaceType4	-0.3850033	2.28e-07
waterDepth	0.0015744	2.10e-05
WSF5	0.0090403	< 2e-16
PRCP	0.0191977	5.89e-08

Table 4.43: Parameters and p-values for selected covariates (Texas)

Louisiana (4,430 observations)

Selected variables:

elevatedBuildingIndicator, numberOfFloorsInTheInsuredBuilding, buildingAgeAtLoss, basementEnclosureCrawlspaceType, waterDepth, PRCP, WSF5

Fitted model

	Estimate	Pr(>|t|)
(Intercept)	10.8454166	< 2e-16
elevatedBuildingIndicator1	-0.4388428	< 2e-16
numberOfFloorsInTheInsuredBuilding	-0.1189752	1.04e-09

buildingAgeAtLoss	-0.0013453	0.0510
basementEnclosureCrawlspaceType1	-0.6276191	< 2e-16
basementEnclosureCrawlspaceType2	-0.4821771	< 2e-16
basementEnclosureCrawlspaceType4	-0.3478725	0.0115
waterDepth	0.0119847	< 2e-16
WSF5	0.0003836	0.7273
PRCP	-0.0615763	< 2e-16

Table 4.44: Parameters and p-values for selected covariates (Louisiana)

Mississippi (254 observations)

Selected variables:

elevatedBuildingIndicator, WSF5, basementEnclosureCrawlspaceType, PRCP, occupancyType

Fitted model

	Estimate	Pr(>|t|)
(Intercept)	11.286321	< 2e-16
basementEnclosureCrawlspaceType1	-0.681720	0.002604
basementEnclosureCrawlspaceType2	-0.778948	0.000211
basementEnclosureCrawlspaceType4	-1.499902	0.012714
elevatedBuildingIndicator1	-0.118845	0.526583
waterDepth	0.004901	0.206897
WSF5	0.030761	0.357330
PRCP	-0.026803	2.16e-06

Table 4.45: Parameters and p-values for selected covariates (Mississippi)

Table 4.43, Table 4.44 and *Table 4.45* present the parameter estimates and *p-values* for variables selected through forward selection in Texas, Louisiana, and Mississippi, respectively.

4.2.2. Model Validation

The Root Mean Square Error (RMSE) for the flood prediction model GLM is 27,115.63 for Texas, 26,488.38 for Louisiana, and 23,065.48 for Mississippi.

4.2.3. Loss Estimation

Texas

	Building 1	Building 2	Building 3
elevatedBuildingIndicator	0	0	0
numberOfFloorsInTheInsuredBuilding	2	3	2
buildingAgeAtLoss	20	15	10
basementEnclosureCrawlspaceType	1	0	0
waterDepth	1.8	1.8	1.8
WSF5	80	80	80
PRCP	3	3	3

Table 4.46: Input Scenarios for GLM Predictions for Buildings in Texas

Louisiana

	Building 1	Building 2	Building 3

elevatedBuildingIndicator	0	0	0
numberOfFloorsInTheInsuredBuilding	2	3	2
buildingAgeAtLoss	20	15	10
basementEnclosureCrawlspaceType	1	0	0
waterDepth	1.8	1.8	1.8
WSF5	80	80	80
PRCP	3	3	3

Table 4.47: Input Scenarios for GLM Predictions for Buildings in Louisiana

Mississippi

	Building 1	Building 2	Building 3
elevatedBuildingIndicator	0	0	0
basementEnclosureCrawlspaceType	1	0	0
waterDepth	1.8	1.8	1.8
WSF5	80	80	80
PRCP	3	3	3

Table 4.48: Input Scenarios for GLM Predictions for Buildings in Mississippi

States	Building 1	Building 2	Building 3	Average
Texas	41,642.53	58,322.23	57,457.08	52,473.95
Louisiana	18,406.67	30,817.53	34,945.35	28,056.52
Mississippi	5,226.345	10,333.926	10,333.926	8,631.40

Table 4.49: Predicted Losses for Buildings

Table 4.46, Table 4. 47 and Table 4. 48 indicate the building characteristics and *Table 4.49* are their loss estimation.

4.3. Annual Expected Losses Calculation

Frequencies are from Table 4.42, and we use the average predicted losses for buildings are from Table 4.49. Combining these, using Equation 3.11, the estimated annual expected losses for wind-related flood events are shown in *Table 4.50*. This table provides an estimation of the financial impact that wind-related flood events are likely to have on buildings in different states on an annual basis.

	Frequency	Average Predicted Loss	Expected Loss
Texas	0.0220	52473.95	1154.427
Louisiana	0.0008	28056.52	22.445
Mississippi	0.0017	8,631.40	14.673

Table 4.50: Annual Expected Loss

5. Discussion

5.1. Interpretation of Key Findings

The thesis focuses on developing a catastrophe model using statistical and machine learning methodologies, particularly the Generalised Linear Model (GLM), to predict losses from extreme wind and flood events. The study is centred on the US Gulf Coast, encompassing Texas, Louisiana, and Mississippi. The key findings are as follows.

5.1.1 Hazard Model Results

The Extreme Value Theory and copula analysis reveal that the Gulf Coast states experience significant variability in extreme weather events, particularly rainfall, wind speed, and flood levels.

Both GEV and GPD analyses show a similar pattern in each variable between these three states.

GEV or GPD analysis shows that the scale parameter increases as one moves from Texas to Mississippi, suggesting greater variability in the intensity of extreme rainfall events. Mississippi has the highest variability, and Texas shows a pronounced heavy-tail behaviour, indicating a higher likelihood of extreme rainfall events.

Texas has the most variable extreme wind speeds and the highest shape parameter, indicating a strong vulnerability to extreme wind events. Louisiana has lower variability and a lighter tail, with Mississippi falling between Texas and Louisiana regarding the potential for extreme wind events.

Extreme flood levels across the states, with Texas showing slightly higher variability. All states have moderately heavy-tailed distributions, with Texas and Louisiana having a slightly higher likelihood of experiencing very extreme flood events compared to Mississippi.

Texas exhibits the highest variability and the strongest tendency towards extreme wind speeds, as seen in its consistently high scale and shape parameters across both GEV and

GPD analyses. This suggests a higher susceptibility to severe wind events. Mississippi generally experiences high variable and intense extreme rainfall events, as indicated by its high-scale parameters in both GEV and GPD analyses. However, its negative shape parameter in the GEV analysis for rainfall suggests a limit to how extreme these events can get. Louisiana shows less variability in rainfall and wind speed extremes than Texas and Mississippi, with a more moderate tail behaviour, meaning less frequent and less severe extreme events.

All three states exhibit heavy-tailed distributions for flood levels, but Louisiana and Texas show a slightly higher propensity for extreme flood events than Mississippi.

5.1.2 Vulnerability Model Results:

The GLM-based vulnerability model highlighted the role of building characteristics and environmental factors in determining the extent of flood damage. The model is able to predict financial losses associated with specific hazard scenarios. Integrating financial loss data into the vulnerability model marks a critical advancement over traditional models, which often focus solely on physical vulnerability. From the GLM parameters, windspeed and flood levels positively correlate with the final losses. However, the parameters for rainfall levels in Louisiana and Mississippi show a negative relationship with that. This might be because the observation to train the vulnerability model in Louisiana and Mississippi is less than in Texas, leading to an estimated error.

Integrating the hazard and vulnerability models provided a tool for predicting potential losses under various extreme weather scenarios. The combined model, validated through cross-validation techniques, demonstrated predictive accuracy, particularly in assessing the financial impact on building infrastructure.

5.1.3 Comparison with Existing Information

The findings align with established theories and research in several ways:

The study's results corroborate existing literature on the intensification of extreme weather events due to climate change. For example, the increase in extreme rainfall and wind speeds in the Gulf Coast states is consistent with the Intergovernmental Panel on Climate Change (IPCC) findings, which have documented the increasing frequency and intensity of such events globally. The heavy-tailed distribution of flood levels further supports the notion that climate change leads to more severe and unpredictable natural disasters.

Also, it is noticeable that the average flood insurance cost in Texas is $678. Comparing this number with the annual expected losses in this paper in Texas $1154)(Fitzpatrick, 2022), the expected flood loss is higher than the insurance cost. However, this is the average cost without mentioning the structure of the insurance product. The higher cost in real life, while lower in the proposed model's prediction, which might cause by the number that comes from the proposed model is the average loss from hurricane-related events, but the insurance policy might only cover the floods from rainfall events and not include heavy rainfall bring by hurricanes in their products.

5.2. Implications for Flood Risk Assessment

One of the most significant contributions of this thesis is integrating financial loss data into the vulnerability model. Traditional models often focus on physical damage without adequately considering the financial implications. Incorporating financial losses with the generalised linear model (GLM) in vulnerability assessment, particularly its application to predicting financial losses, showcases originality in applying machine learning techniques in flood risk analysis. The thesis provides a comprehensive tool for disaster risk management, offering valuable insights for the insurance and reinsurance industries. This approach reflects a growing recognition in the field that economic resilience is as crucial as physical resilience in mitigating the impacts of natural disasters.

The proposed catastrophe modelling framework incorporates both hazard and vulnerability models, with the final output—annual expected loss—being derived as the product of the results from these models. Notably, these two components within the framework can operate independently. For instance, the model can calculate the probability of exceeding a certain threshold annually if users are interested in the joint exceedance probability, or it can estimate the potential loss for a specific building or portfolio of buildings given a particular event intensity.

This research shows an example of how knowledge is created and interpreted within the field of disaster risk management. It proposes a catastrophe modelling framework that builds on existing theories and models and pushes the boundaries by integrating advanced statistical methods and machine learning techniques. This approach addresses the complexities inherent in multi-hazard scenarios, offering a more nuanced and powerful tool for managing disaster risk.

5.3. Research Limitations, Future Directions and Recommendations

The thesis is limited to the US Gulf Coast (Texas, Louisiana and Mississippi), so the findings may not directly apply to other regions with different geographic or climatic conditions. This limits the generalisability of the model. Future research could expand the model to other regions with different climatic and geographic conditions. This would test the model's adaptability and robustness across various environments.

After this, this catastrophe framework also has the limitation that the codes might only apply to one geographical location because of the complexity of the copula fitting process and the variation between choosing different types of copula in prediction. This limit requires users to modify the codes while analysing other geographical locations. For example, models between Texas, Louisiana and Mississippi use the main framework and methodology but slightly different codes to choose the best fit copula and give the joint probability. This indicates that further research could find out as many variations as possible when analysing/fitting more data from different geographical locations. This can

make the model run more smoothly and be user-friendly. Also, this proposed catastrophe modelling framework is only limited to fit four types of copula (Gumbel, Clayton, Frank and t copula). Continuing fitting the copula model to the data can increase the accuracy of probability estimations, although this work has the potential to lead the model to become more complex since it includes more variations and makes the process automatic.

The model may simplify the interactions between different hazards (e.g., wind and flood), which could lead to underestimating risks in scenarios where these interactions are more complex. This underestimation is particularly shown in calculating the annual joint probability in Louisiana. One of the reasons for this might be using the 3-dimensional copula analysis. Further research can include fitting the variables pairwise in a 2-dimensional copula. This might increase the computational complexity but can increase the accuracy of analysing the dependency structure between each pair of variables.

In addition, while this thesis focuses on wind-related flood hazards, future research could extend the model to include other types of natural hazards, such as earthquakes or tsunamis, and examine the interactions between multiple hazards (e.g., how an earthquake might exacerbate flooding).

Another limitation of this model is that while fitting the GPD, the user is required to estimate/select the thresholds manually. This process might require the user to have statistical modelling experience, extreme value theory knowledge, and an understanding of the local environment. One possible recommendation is to conduct workshops and training sessions for stakeholders, including government agencies, local authorities, insurance companies, and urban planners, to educate them on the use and interpretation of the model. This helps ensure that the model is applied effectively in decision-making processes.

A further limitation in the research is that while we fitted the data to GEV and GPD, we did not check if a non-stationary GEV or GPD would perform better. Further research can include fitting the non-stationary GEV and GPD to examine this. Because of climate change, the distribution parameters will likely change over time. If indeed the non-stationary model fitted performs better, using copula analysis will become more complex.

The complexity can come from modelling and accounting for non-stationarity, such as transforming the data to achieve stationarity before applying the copula analysis or using dynamic copulas that can handle non-stationary dependencies. However, integrating climate change adaptation strategies into the model's framework can help users respond to current risks and prepare for future changes in hazard patterns.

Regarding the data limitation in this research, the meteorology data used in this study are data from airport recodes, which are far from the coastal line and it is rare to measure flood levels in this area. If we use the observation data from inland, it is very likely to underestimate the intensity of the extreme events, particularly in measuring and modelling the windspeed. This might be the second reason that explains why annual joint probability in Louisiana is very low. Although the water depth (flood level) data calculated from the insurance claims records fit the distributions in extreme value theory well, it is not recorded from direct measurement in the natural environment. Also, the vulnerability model in Mississippi shows a potential underestimation of the financial losses. This might be because there are only 254 observations to train the model.

The model's accuracy and reliability heavily depend on the quality and availability of data. If possible, further research can use direct flood level data from the environmental agent to obtain more reliable results. The model's predictions may be less accurate in regions where data is scarce or less reliable. For instance, the flood level data used in this paper is a calculated average water depth from the insurance record, not observation data from the local authority. To address data limitations, one of the recommendations for continuing to work on the observation data is for governmental and private/commercial organisations to collaborate to improve the collection and sharing of high-quality meteorological data. This can also include efforts to standardise data collection methods across different regions. Also, engage with the insurance and reinsurance industries to refine the model further based on their needs and feedback. This collaboration also provides opportunities for real-world testing and validation of the model in live environments. Collaboration with industry would be beneficial in increasing the accuracy of estimating financial losses, which can lead to preparing the funding/reserve for flood recovery precisely. On the other hand,

it is also worth using the model-simulated data generated by climate models, such as data generated in Coupled Model Intercomparison Project Phase 6 (CMIP6), to fit the proposed multi-hazard framework. Using climate model simulated data can solve the data limitations in the hazard model. For example, it does not constrain the location, and the data points can be more precise than the observation data point.

One notable point is the increasing popularity and precision of advanced machine learning algorithms across various quantitative fields globally. The vulnerability component discussed in this paper could potentially achieve better statistical performance by using these algorithms to train the same dataset. Future research could explore the enhancement of vulnerability models by incorporating ensemble learning algorithms, such as the Gradient Boosting Machine (GBM), alongside the Generalised Linear Model (GLM) to facilitate a comprehensive comparison. While deep learning algorithms undeniably have the potential to perform better in prediction statistically, the need for explainability and the consideration of the model's users suggest that explainable and supervised machine learning algorithms like GLM and GBM might offer a more balanced and effective overall performance.

6. Conclusion

This study has successfully developed a flood risk assessment model tailored to the unique challenges of the US Gulf Coast, with a specific focus on Texas, Louisiana, and Mississippi. Given the increasing flood risks due to climate change and urbanisation, accurately predicting the frequency of extreme flood events and the expected financial losses from these events is crucial. Misleading or inaccurate models can have serious consequences, leading policymakers and authorities to make less ideal decisions that exacerbate the impact of catastrophic flooding events. Therefore, the reliability and clarity of flood catastrophe modelling tools are essential.

The potential application of this model extends beyond financial loss predictions. For policymakers and urban planners in Texas, the hazard model offers a valuable tool for long-term disaster risk reduction strategies by advising authorities on the exceedance probability of flood events. By providing clear and reliable predictions, the complete model can inform infrastructure investments, emergency preparedness plans, and insurance policies, ultimately contributing to the resilience of communities against future flood events. In addition, the hazard model is able to calculate the level of return and probability of overruns at different thresholds, which provides decision-makers with the data they need to make informed decisions on resource allocation and flood prevention efforts.

To ensure the model's practical application in managing flood risks, it is recommended that workshops and training sessions be conducted for stakeholders, including government agencies, local authorities, and insurance companies, because of the model's complexity, particularly in the manual selection of thresholds for the Generalised Pareto Distribution (GPD). These sessions would ensure that the model is applied effectively in decision-making processes and that users fully understand its limitations and capabilities.

In conclusion, this research contributes to non-commercial flood risk modelling for the US Gulf Coast, with the potential for broader applications. By addressing the identified limitations and incorporating future advancements, the proposed modelling framework could become a powerful tool for global flood risk assessment, helping communities and policymakers worldwide to better prepare for and respond to extreme weather events.

References

[1] United Nations Department of Economic and Social Affairs, Population Division. (2022). World Population Prospects 2022 Summary of Results. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf (Access date: 29/01/2024)

[2] Gu, D. (2019). Population Division Exposure and vulnerability to natural disasters for world's cities*. https://www.un.org/en/development/desa/population/publications/pdf/technical/TP2019-4.pdf

[3] Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., & Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. Nature, 596(7870), 80–86. https://doi.org/10.1038/s41586-021-03695-w

[6] Scaife A. A., Comer, R., Dunstone, N., Fereday, D., Folland, C., Good, E., Gordon, M., Hermanson, L., Ineson, S., Karpechko, A., Knight, J., MacLachlan, C., Maidens, A., Peterson, K. A., Smith, D., Slingo, J. & Walker, B. (2017) Predictability of European Winter 2015/16. Atmospheric Science Letters, 18, 38-44. https://doi.org/10.1002/asl.721

[7] Fung, F., Bett, P., Maisey, P., Lowe, J., McSweeney, C., Mitchell, J. F. B., Murphy, J., Rostron, J., Sexton, D. & Yamazaki, K. (2018) UKCP18 Factsheet: Wind. Met Office Hadley Centre, Exeter. https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-fact-sheet-wind_march21.pdf

[8] Lowe, J. A., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D., Clark, R., Eagle, K., Edwards, T., Fosser, G., Fung, F., Gohar, L., Good, P., Gregory, J., Harris, G., Howard, T., Kaye, N., Kendon, E., Krijnen, J., Maisey, P., McDonald, R., McInnes, R., McSweeney, C., Mitchell, J. F.B., Murphy, J., Palmer, M., Roberts, C., Rostron, J., Sexton, D., Thornton, H., Tinker, J., Tucker, S., Yamazaki, K. & Belcher, S. (2018) UKCP18 science overview report. Met Office Hadley Centre, Exeter.

https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf

[9] McDonald, R. E. (2011) Understanding the impact of climate change on northern hemisphere extra-tropical cyclones. Climate Dynamics, 37, 1399-1425. https://doi.org/10.1007/s00382-010-0916-x

[10] Jane, R., Cadavid, L., Jayantha Obeysekera, & Wahl, T. (2020). Multivariate statistical modelling of the drivers of compound flood events in South Florida. https://doi.org/10.5194/nhess-2020-82

[11] Moya-Álvarez, A., Gálvez, J., Holguín, A., Estevan, R., Kumar, S., Villalobos, E., Martínez-Castro, D., & Silva, Y. (2018). Extreme Rainfall Forecast with the WRF-ARW Model in the Central Andes of Peru. Atmosphere, 9(9), 362. https://doi.org/10.3390/atmos9090362

[12] Lin, N., Smith, J. A., Villarini, G., Marchok, T. P., & Baeck, M. L. (2010). Modeling Extreme Rainfall, Winds, and Surge from Hurricane Isabel (2003). Weather and Forecasting, 25(5), 1342–1361. https://doi.org/10.1175/2010waf2222349.1

[13] Tian, J., Liu, J., Wang, J., Li, C., Yu, F., & Chu, Z. (2017). A spatio-temporal evaluation of the WRF physical parameterisations for numerical rainfall simulation in semi-humid and semi-arid catchments of Northern China. Atmospheric Research, 191, 141–155. https://doi.org/10.1016/j.atmosres.2017.03.012

[14] Johansson, B., & Chen, D. (2003). The influence of wind and topography on precipitation distribution in Sweden: Statistical Analysis and modelling. International Journal of Climatology, 23(12), 1523–1535. https://doi.org/10.1002/joc.951

[15] Antzoulatos, G., Kouloglou, I.-O., Bakratsas, M., et al. (2022). Flood Hazard and Risk Mapping by Applying an Explainable Machine Learning Framework Using Satellite Imagery and GIS Data. https://doi.org/10.3390/su14063251

[16] Maina, S. C., Mwigereri, D., Weyn, J., Mackey, L., & Ochieng, M. (2023). Evaluation of Dependency Structure for Multivariate Weather Predictors Using Copulas. ACM Journal on Computing and Sustainable Societies, 1(2), 1–23. https://doi.org/10.1145/3616384

[17] Friederichs, P., Wahl, S., & Buschow, S. (2018). Postprocessing for extreme events. Statistical Postprocessing of Ensemble Forecasts, 127–154. https://doi.org/10.1016/b978-0-12-812372-0.00005-4

[18] Tabari, H. (2021). Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. Journal of Hydrology, 593, 125932. https://doi.org/10.1016/j.jhydrol.2020.125932

[19] Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-70816-2

[20] Madsen, H., Lawrence, D., Lang, M., Martinkova, M., & Kjeldsen, T. R. (2014). Review of Trend Analysis and climate change projections of extreme precipitation and floods in Europe. Journal of Hydrology, 519, 3634–3650. https://doi.org/10.1016/j.jhydrol.2014.11.003

[21] Poulomi Ganguli; M. Janga Reddy; “Probabilistic Assessment of Flood Risks Using Trivariate Copulas”, THEORETICAL AND APPLIED CLIMATOLOGY, 2012.

[22] Thorsten Schmidt (2006) "Coping with Copulas", http://archiv.stochastik.uni-freiburg.de/homepages/schmidt/publications/TSchmidt_Copulas.pdf

[23] Low, R. K., Alcock, J., Faff, R., & Brailsford, T. (2013). Canonical vine copulas in the context of modern portfolio management: Are they worth it? Journal of Banking & Finance, 37(8), 3085–3099. https://doi.org/10.1016/j.jbankfin.2013.02.036

[24] Demarta, S., & McNeil, A. J. (2007). The t copula and related copulas. International Statistical Review, 73(1), 111–129. https://doi.org/10.1111/j.1751-5823.2005.tb00254.x

[25] Jobst, D., Möller, A., & Groß, J. (2023). D-vine-copula-based postprocessing of wind speed ensemble forecasts. https://doi.org/10.1002/qj.4521

[26] Molnar, C., Casalicchio, G. and Bischl, B. (2020) “Interpretable Machine Learning -- A Brief History, State-of-the-Art and Challenges,” arXiv (Cornell University) [Preprint]. Available at: https://arxiv.org/pdf/2010.09337.pdf.

[30] Blake, E., & Zelinsky, D. (2018). HURRICANE HARVEY. https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf

[31] Keim, B. D., & Muller, R. A. (2009). Hurricanes of the Gulf of Mexico. In Google Books. LSU Press. https://books.google.co.uk/books?id=7B9fDwAAQBAJ&printsec=frontcover#v=onepage&q&f=false

[32] Eilander, D., Couasnon, A., Leijnse, T., Ikeuchi, H., Yamazaki, D., Muis, S., Dullaart, J., Haag, A., Winsemius, H. C., & Ward, P. J. (2023). A globally applicable framework for compound flood hazard modeling. Natural Hazards and Earth System Sciences, 23(2), 823–846. https://doi.org/10.5194/nhess-23-823-2023

[33] Couasnon, A., Eilander, D., Muis, S., Veldkamp, T. I. E., Haigh, I. D., Wahl, T., Winsemius, H. C., & Ward, P. J. (2020). Measuring compound flood potential from river discharge and storm surge extremes at the global scale. Natural Hazards and Earth System Sciences, 20(2), 489–504. https://doi.org/10.5194/nhess-20-489-2020

[34] Wahl, T., Jain, S., Bender, J., Meyers, S. D., & Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. Nature Climate Change, 5(12), 1093–1097. https://doi.org/10.1038/nclimate2736

[35] Thieken, A. H., Kienzler, S., Kreibich, H., Kuhlicke, C., Kunz, M., Mühr, B., Müller, M., Otto, A., Petrow, T., Pisi, S., & Schröter, K. (2016). Review of the flood risk management system in Germany after the major flood in 2013. Ecology and Society, 21(2). https://doi.org/10.5751/es-08547-210251

[36] Fitzpatrick, M. (2022, August 5). *How Much Is Flood Insurance in Texas? Average Flood Insurance Cost and Coverage*. MoneyGeek.com. https://www.moneygeek.com/insurance/flood/texas-flood-insurance-costs/

[37] Adelekan, I. O. (2010). Vulnerability assessment of an urban flood in Nigeria: Abeokuta flood 2007. *Natural Hazards*, *56*(1), 215–231. https://doi.org/10.1007/s11069-010-9564-z

[38] Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., & Thandi, A. (2007). *A Practitioner's Guide to Generalized Linear Models Color version*. https://www.casact.org/sites/default/files/database/dpp_dpp04_04dpp1.pdf

[39] Bakkensen, L. A., Fox-Lent, C., Read, L. K., & Linkov, I. (2016). Validating Resilience and Vulnerability Indices in the Context of Natural Disasters. *Risk Analysis*, *37*(5), 982–1004. https://doi.org/10.1111/risa.12677

[40] Bezak, N., Mikoš, M., & Šraj, M. (2014). Trivariate Frequency Analyses of Peak Discharge, Hydrograph Volume and Suspended Sediment Concentration Data Using Copulas. *Water Resources Management*, *28*(8), 2195–2212. https://doi.org/10.1007/s11269-014-0606-2

[41] Birkland, T. A. (2006). *Lessons of Disaster: Policy Change after Catastrophic Events*. https://doi.org/10.1353/book13054

[42]Cutter, S. L. (1996). Vulnerability to environmental hazards. *Progress in Human Geography*, *20*(4), 529–539. https://doi.org/10.1177/030913259602000407

[43] Cutter, S. L. (2003). The Vulnerability of Science and the Science of Vulnerability. *Annals of the Association of American Geographers*, *93*(1), 1–12. https://doi.org/10.1111/1467-8306.93101

[44] Genest, C., Kilani Ghoudi, & Rivest, L.-P. (1995). *A semiparametric estimation procedure of dependence parameters in multivariate families of distributions*. *82*(3), 543–552. https://doi.org/10.1093/biomet/82.3.543

[45] Goldburd, M., Khare, A., Tevet, D., & Guller, D. (2020). *Casualty Actuarial Society CAS MONOGRAPH SERIES NUMBER 5 Second Edition GENERALIZED LINEAR MODELS FOR INSURANCE RATING Second Edition*. https://www.casact.org/sites/default/files/2021-01/05-Goldburd-Khare-Tevet.pdf

[46] Grenier, R., Sousounis, J., Schneyer, D., & Raizman. (n.d.). *Quantifying the Impact from Climate Change on U.S. Hurricane Risk*. https://www.air-worldwide.com/siteassets/Publications/White-Papers/documents/air_climatechange_us_hurricane_whitepaper.pdf

[47] Hufschmidt, G. (2011). A comparative analysis of several vulnerability concepts. *Natural Hazards*, *58*(2), 621–643. https://doi.org/10.1007/s11069-011-9823-7

[48] Khosravi, K., Pourghasemi, H. R., Chapi, K., & Bahri, M. (2016). Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. *Environmental Monitoring and Assessment*, *188*(12). https://doi.org/10.1007/s10661-016-5665-9

[49] Loganathan, G. V., Kuo, C. Y., & J. Yannaccone. (1987). Joint Probability Distribution of Streamflows and Tides in Estuaries. *Hydrology Research*, *18*(4-5), 237–246. https://doi.org/10.2166/nh.1987.0017

[50] Mentzafou, A., Markogianni, V., & Dimitriou, E. (2016). The Use of Geospatial Technologies in Flood Hazard Mapping and Assessment: Case Study from River Evros. *Pure and Applied Geophysics*, *174*(2), 679–700. https://doi.org/10.1007/s00024-016-1433-6

[51] Morrow, B. H. (1999). Identifying and Mapping Community Vulnerability. Disasters, 23(1), 1–18. https://doi.org/10.1111/1467-7717.00102

[52] Nelsen, R. B. (2011). An introduction to copulas. Springer.

[53] Nkwunonwo, U. C., Whitworth, M., & Baily, B. (2020). A review of the current status of flood modelling for urban flood risk management in the developing countries. *Scientific African*, *7*, e00269. https://doi.org/10.1016/j.sciaf.2020.e00269

[54] Pietro Parodi. (2015). *Pricing in general insurance*. Taylor & Francis.

[55] Poudyal, N. C., Johnson-Gaither, C., Goodrick, S., Bowker, J. M., & Gan, J. (2012). Locating Spatial Variation in the Association Between Wildland Fire Risk and Social Vulnerability Across Six Southern States. Environmental Management, 49(3), 623–635. https://doi.org/10.1007/s00267-011-9796-z

[56] Scheuer, S., Haase, D., & Meyer, V. (2010). Exploring multicriteria flood vulnerability by integrating economic, social and ecological dimensions of flood risk and coping

capacity: from a starting point view towards an end point view of vulnerability. Natural Hazards, 58(2), 731–751. https://doi.org/10.1007/s11069-010-9666-7

[57] Sharifi, A. (2016). A critical review of selected tools for assessing community resilience. *Ecological Indicators*, *69*, 629–647. https://doi.org/10.1016/j.ecolind.2016.05.023

[58] Towe, R., Tawn, J., Lamb, R., Sherlock, C., & Liu, Y. (2016). Improving statistical models for flood risk assessment. *E3S Web of Conferences*, *7*, 01011. https://doi.org/10.1051/e3sconf/20160701011

[59] Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Kasperson, J. X., Luers, A., Martello, M. L., Polsky, C., Pulsipher, A., & Schiller, A. (2003). A framework for vulnerability analysis in sustainability science. *Proceedings of the National Academy of Sciences*, *100*(14), 8074–8079. https://doi.org/10.1073/pnas.1231335100

[60] Coles, S., & Casson, E. (1998). Extreme value modelling of hurricane wind speeds. Structural Safety, 20(3), 283–296. https://doi.org/10.1016/s0167-4730(98)00015-0

[61] Davison, A. C., & Smith, R. L. (1990). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

[62] Gallina, V., Torresan, S., Critto, A., Sperotto, A., Glade, T., & Marcomini, A. (2016). A review of multi-risk methodologies for natural hazards: Consequences and challenges for a climate change impact assessment. *Journal of Environmental Management*, *168*, 123–132. https://doi.org/10.1016/j.jenvman.2015.11.011

[63] Johnson, K., Depietri, Y., & Breil, M. (2016). Multi-hazard risk assessment of two Hong Kong districts. International Journal of Disaster Risk Reduction, 19, 311–323. https://doi.org/10.1016/j.ijdrr.2016.08.023

[64] Ming, X., Liang, Q., Dawson, R., Xia, X., & Hou, J. (2022). A quantitative multi-hazard risk assessment framework for compound flooding considering hazard inter-

dependencies and interactions. *Journal of Hydrology*, *607*, 127477. https://doi.org/10.1016/j.jhydrol.2022.127477

[65] Picard, R. R., & Cook, R. D. (1984). Cross-Validation of Regression Models. *Journal of the American Statistical Association*, *79*(387), 575–583. https://doi.org/10.2307/2288403

[66] Hallegatte, S., Green, C., Nicholls, R. J., & Corfee-Morlot, J. (2013). Future flood losses in major coastal cities. *Nature Climate Change*, *3*(9), 802–806. https://doi.org/10.1038/nclimate1979

[67] Aerts, J. C. J. H., Botzen, W. J., Clarke, K. C., Cutter, S. L., Hall, J. W., Merz, B., Michel-Kerjan, E., Mysiak, J., Surminski, S., & Kunreuther, H. (2018). Integrating human behaviour dynamics into flood disaster risk assessment. *Nature Climate Change*, *8*(3), 193–199. https://doi.org/10.1038/s41558-018-0085-1

[68] Bachner, G., Knittel, N., Poledna, S., Hochrainer-Stigler, S., & Reiter, K. (2023). Revealing indirect risks in complex socioeconomic systems: A highly detailed multi-model analysis of flood events in Austria. *Risk Analysis*, *44*(1), 229–243. https://doi.org/10.1111/risa.14144

[69] Botzen, W. J. W., Aerts, J. C. J. H., & van den Bergh, J. C. J. M. (2009). Dependence of flood risk perceptions on socioeconomic and objective risk factors. *Water Resources Research*, *45*(10). https://doi.org/10.1029/2009wr007743

[70] *Climate Prediction Center - Atlantic Hurricane Outlook*. (2024). Noaa.gov. https://www.cpc.ncep.noaa.gov/products/outlooks/hurricane2024/May/hurricane.shtml

[71] de Ruig, L. T., Haer, T., de Moel, H., Brody, S. D., Botzen, W. J. W., Czajkowski, J., & Aerts, J. C. J. H. (2022). Climate-proofing the National Flood Insurance Program. *Nature Climate Change*. https://doi.org/10.1038/s41558-022-01502-6

[72] Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., & Zhang, X.

(2018). Future climate risk from compound events. *Nature Climate Change*, *8*(6), 469–477. https://doi.org/10.1038/s41558-018-0156-3

[73] IPCC. (2021). *The Physical Science Basis Climate Change 2021 Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. https://report.ipcc.ch/ar6/wg1/IPCC_AR6_WGI_FullReport.pdf

[74] Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C. J. H., Mechler, R., Botzen, W. J. W., Bouwer, L. M., Pflug, G., Rojas, R., & Ward, P. J. (2014). Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, *4*(4), 264–268. https://doi.org/10.1038/nclimate2124

[75] Jongman, B., Winsemius, H. C., Aerts, J. C. J. H., Perez, E. C. de, Aalst, M. K. van, Kron, W., & Ward, P. J. (2015). Declining vulnerability to river floods and the global benefits of adaptation. *Proceedings of the National Academy of Sciences*, *112*(18), E2271–E2280. https://doi.org/10.1073/pnas.1414439112

[76] Kappes, M. S., Keiler, M., von Elverfeldt, K., & Glade, T. (2012). Challenges of analyzing multi-hazard risk: a review. *Natural Hazards*, *64*(2), 1925–1958. https://doi.org/10.1007/s11069-012-0294-2

[77] Kunreuther, H. C., & Michel-Kerjan, E. O. (2009). *At War with the Weather*. https://doi.org/10.7551/mitpress/9780262012829.001.0001

[78] Lamond, J., & Penning-Rowsell, E. (2014). The robustness of flood insurance regimes given changing risk resulting from climate change. *Climate Risk Management*, *2*, 1–10. https://doi.org/10.1016/j.crm.2014.03.001

[79] Li, F., Liu, H., Chen, X., & Yu, D. (2019). Trivariate Copula Based Evaluation Model of Water Accessibility. *Water Resources Management*, *33*(9), 3211–3225. https://doi.org/10.1007/s11269-019-02292-x

[80] Mechler, R., Bouwer, L. M., Linnerooth-Bayer, J., Hochrainer-Stigler, S., Aerts, J. C. J. H., Surminski, S., & Williges, K. (2014). Managing unnatural disaster risk from climate extremes. *Nature Climate Change*, *4*(4), 235–237. https://doi.org/10.1038/nclimate2137

[81] Michel-Kerjan, E. (2012). How resilient is your country? *Nature*, *491*(7425), 497–497. https://doi.org/10.1038/491497a

[82] Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B. (2020). General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2007.04131

[83] Shiau, J. T., & Modarres, R. (2009). Copula-based drought severity-duration-frequency analysis in Iran. *Meteorological Applications*, *16*(4), 481–489. https://doi.org/10.1002/met.145

[84] Sousounis, P., & Little, C. (2017). *Climate Change Impacts on Extreme Weather*. https://www.air-worldwide.com/SiteAssets/Publications/White-Papers/documents/Climate-Change-Impacts-on-Extreme-Weather

[85] Surminski, S., & Oramas-Dorta, D. (2014). Flood insurance schemes and climate adaptation in developing countries. *International Journal of Disaster Risk Reduction*, *7*, 154–164. https://doi.org/10.1016/j.ijdrr.2013.10.005

[86] U.S. Global Change Research Program. (2017). Fourth National Climate Assessment. *Globalchange.gov*, *2*(4), 1–470. https://nca2018.globalchange.gov/

[87] The Coastline at Risk. (n.d.). AIRWorldwide. Retrieved May 15, 2024, from https://www.air-worldwide.com/Models/Tropical-Cyclone/The-Coastline-at-Risk/

[88] $1 in 1978 → 2024 | Inflation Calculator. (n.d.). Www.in2013dollars.com. Retrieved May 15, 2024, from https://www.in2013dollars.com/us/inflation/1978?amount=1#formulas

[89] FIMA NFIP Redacted Claims - v2 | FEMA.gov. (2023, July 5). Www.fema.gov. https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v2

[90] *NOAA | Climate Data Online (CDO) | National Climatic Data Center (NCDC)*. (n.d.). Www.ncei.noaa.gov. https://www.ncei.noaa.gov/cdo-web/search

Appendix:

Name	Title	Description
causeOfDamage	Cause Of Damage	Categorical variable indicating the cause of damage, with "4" denoting rain-related flood damage.
buildingDamageAmount	Building Damage Amount	The actual cash value amount of damage to a main property in whole dollars.
basementEnclosure CrawlspaceType	Basement Enclosure Crawlspace Type	Categorical - Specifies the type of basement or crawlspace, with values ranging from 0 to 4; 0 - None; 1 - Finished Basement/Enclosure; 2 - Unfinished Basement/Enclosure; 3 - Crawlspace; 4 - Subgrade Crawlspace;
dateOfLoss	Date of Loss	Categorical - The date when water first entered the insured building.
elevatedBuildingIndicator	Elevated Building Indicator	Binary - Indicates whether the building is elevated (Yes = 'true' or '1'; No = 'false' or '0').
ratedFloodZone	Rated Flood Zone	*Categorical* - The NFIP Flood Zone as per the Flood Insurance Rate Map (FIRM).
locationOfContents	Location of Contents	Categorical - Location code of contents: 1- Basement/Enclosure/Crawlspace/Subgrade Crawlspace only; 2 - Basement/Enclosure/Crawlspace/Subgrade Crawlspace and above; 3 - Lowest floor only above ground level (no basement/enclosure/crawlspace/subgrade crawlspace); 4 - Lowest floor above ground level and higher floors (no basement/enclosure/crawlspace/subgrade crawlspace); 5 - Above ground level more than one full floor; 6 - Manufactured (mobile) home or travel trailer on foundation; 7 - Enclosure/Crawlspace and above;
numberOfFloorsInTheInsuredBuilding	Number of Floors in the Insured Building	Categorical - Indicates the number of floors in the insured building, ranging from 1 to 6. 1 = One floor; 2= Two floors; 3 = Three or more floors; 4 = Split-level; 5 = Manufactured (mobile) home or travel trailer on foundation; 6 = Townhouse/Rowhouse with three or more floors (RCBAP Low-rise only);

nonProfitIndicator	Non-Profit Indicator	Binary - Indicates if the building is reported as non-profit (Yes = 'true' or '1'; No = 'false' or '0').
occupancyType	Occupancy Type	Categorical - Code indicating the use and occupancy of the insured structure, like single-family residence or non-residential building. 1=single family residence; 2 = 2 to 4 unit residential building; 3 = residential building with more than 4 units; 4 = Non-residential building; 6 = Non Residential - Business; 11 = Single-family residential building with the exception of a mobile home or a single residential unit within a multi unit building; 12 = A residential non-condo building with 2, 3, or 4 units seeking insurance on all units; 13 = A residential non-condo building with 5 or more units seeking insurance on all units; 14 = Residential mobile/manufactured home; 15 = Residential condo association seeking coverage on a building with one or more units; 16 = Single residential unit within a multi-unit building; 17 = Non-residential mobile/manufactured home; 18 = A non-residential building; 19 = a non-residential unit within a multi-unit building;
originalConstructionDate	Original Construction Date	Categorical - The original date of the construction of the building.
floodWaterDuration	Flood Water Duration	Continuous - Number of hours flood water remained in the insured building
floodproofedIndicator	Floodproofed Indicator	Binary - Indicates if the structure is floodproofed (Yes = 'true' or '1'; No = 'false' or '0').
numberOfUnits	Number of Units	Continuous - The number of residential and non-residential units covered by the policy.
stateOwnedIndicator	State-Owned Indicator	Binary - Indicates if the property is state-owned (Yes = 'true' or '1'; No = 'false' or '0').
waterDepth	Water Depth	Continuous - Depth of flood water inches (in).
state	state	Categorical - The two-character abbreviation of the state where the property is located.
buildingAgeAtLoss	Building Age At Loss	Continuous - The age of the building at the time of loss.
DATE	DATE	Categorical - The specific date of recorded weather data.
PRCP	Rainfall Level	Continuous - Daily precipitation levels in inches (in).

WSF5	Windspeed	Continuous - The fastest 5-second windspeed recorded on that day in Miles Per Hour (mph).

Table A1: Variable Definitions for Assessment Database