Understanding Extreme Rainfall Behavior Using Tail Characterization

Vincent Weng

Introduction

In this memo we investigates the distribution of extreme rainfall by focusing on characterizing the tails of the marginal distribution of precipitation in Miami. Specifically, we aim to assess the lightness/heaviness of the tail, quantifying the frequency and intensity of extreme rainfall events and identifying appropriate methods for modeling their occurrence. Historical precipitation data for Miami was obtained from the National Centers for Environmental Information between 1958 and 2025. After data preprocessing, the dataset contains about 24,000 daily observations, providing a robust foundation for analyzing extreme rainfall patterns. Understanding the extremity of precipitation events is critical for assessing the likelihood and severity of future rainfall, which has direct implications for flood risk management, infrastructure resilience, and climate adaptation planning. Accurately modeling precipitation tails helps anticipate hazards and guide water management and urban planning.

Methods

This study applies extreme value theory using both the Generalized Pareto Distribution and Generalized Extreme Value Distribution to analyze the statistical behavior of extreme rainfall. The GPD approach was implemented by selecting a threshold to define exceedances, with candidate thresholds evaluated using quantile-based methods. Observations surpassing the chosen threshold were extracted, and their exceedances (the amount exceeding the threshold) were used for tail modeling. To characterize the tail behavior, exponential decay, power-law distributions, and the Generalized Pareto distribution were considered. This provides a flexible framework to capture whether it's heavy-tailed or light-tailed. Fitting a regression line was utilized to estimate the shape and scale parameters of the fitted model and then root mean squared error (RMSE) was used to determine the best fit.

Additionally, the Block Maxima approach was applied by segmenting the dataset into yearly blocks and extracting the maximum rainfall value within each year. The Generalized Extreme Value distribution was then fitted to these block maxima to model the probability distribution of extreme rainfall events over fixed time periods. This approach allowed for comparison between the GPD-based exceedance model and the GEV-based block maxima model, providing complementary perspectives on extreme rainfall behavior. The estimated parameters from both methods were used to compute return levels—the rainfall amounts expected to be exceeded once in a given return period. By comparing the shape parameters and fit diagnostics of both GPD and GEV, this study evaluates whether extreme rainfall follows an exponential decay (light-tailed) or power-law (heavy-tailed).

Results

The GPD analysis was conducted on different subsets of the data, corresponding to the top 25% (7,039 observations), top 10% (2,816 observations), top 5% (1,408 observations), and top 1% (282 observations). For the power-law fitting, the RMSE values were 1.426, 1.727, 2.828, and 4.189 for these subsets, respectively (Figure 1). For the exponential fitting, the RMSE values were 1.267, 1.566, 2.800, and 4.189, respectively (Figure 2).The GPD fit produced a shape parameter of 0.022 and a scale parameter of 15.07. The estimated return levels from the GPD model were 99.33 mm for the 1-year return period, 203.21 mm for the 10-year return period, 375.39 mm for the 100-year return period, and 660.79 mm for the 1000-year return period**)**.

The GEV fit to the block maxima yielded a shape parameter of -0.208, indicating a bounded tail. The estimated return levels for extreme rainfall events were 192.25 mm for the 10-year return period, 357.42 mm for the 100-year return period, and 620.64 mm for the 1000-year return period. A comparison of return levels between the GPD and GEV models was also plotted to visualize their closeness in alignment across different return periods (Figure 3).

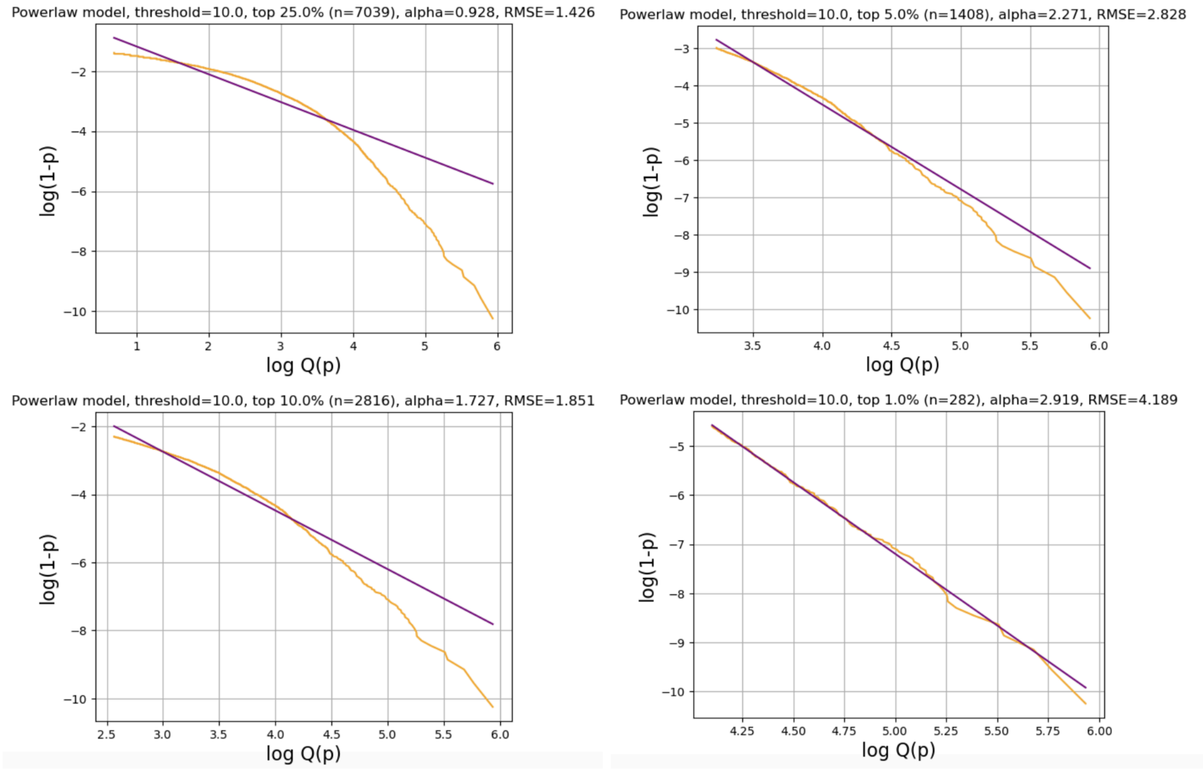
Interpretation

The RMSE values indicate that the exponential model provides a better fit to moderate rainfall extremes than the power-law model, suggesting that exceedances above a threshold exhibit an approximately exponential decay. This aligns with the GPD results, which captured the tail behavior but favored an exponential-like distribution rather than a strongly heavy-tailed power-law. However, the GEV model, which was fitted to annual maxima, produced a shape parameter indicative of a bounded tail, meaning that extreme yearly rainfall events have a natural upper limit rather than following a power-law decay. This suggests that while individual threshold exceedances in GPD approximate an exponential-like decay, the most extreme yearly maxima modeled by GEV are bounded rather than indefinitely heavy-tailed.

The return level comparison between GEV and GPD in the graph further supports this distinction. The two models produce similar return level estimates across different return periods, indicating that they largely agree on the likelihood of extreme events. However, the GPD return levels are slightly higher for longer return periods, which may be attributed to its sensitivity to frequent exceedances. This suggests that, even though GPD favors an exponential decay in moderate extremes, it still captures more tail heaviness than a strictly bounded distribution. Meanwhile, GEV's focus on annual maxima leads to a more constrained representation of long-term tail behavior, reinforcing the bounded nature of rare but severe rainfall events.

These results highlight an important nuance in extreme rainfall modeling: GPD, when fitted with an exponential-like decay, effectively characterizes moderate extremes but may slightly overestimate the probability of the most severe events. GEV, by capturing annual maxima, emphasizes long-term extreme event probabilities but suggests that the most extreme rainfall events are ultimately bounded, rather than following a power-law or even a fully exponential decay. However, one key consideration is the uncertainty in describing the marginal distribution of the tail. While both models provide reasonable fits, differences in their approaches could influence tail estimates, especially for extreme return periods. The reliance on block maxima in GEV versus threshold exceedances in GPD introduces potential biases depending on data availability and model assumptions. Additionally, variations in the choice of threshold for GPD or block size for GEV could lead to different conclusions about tail heaviness.

Figures

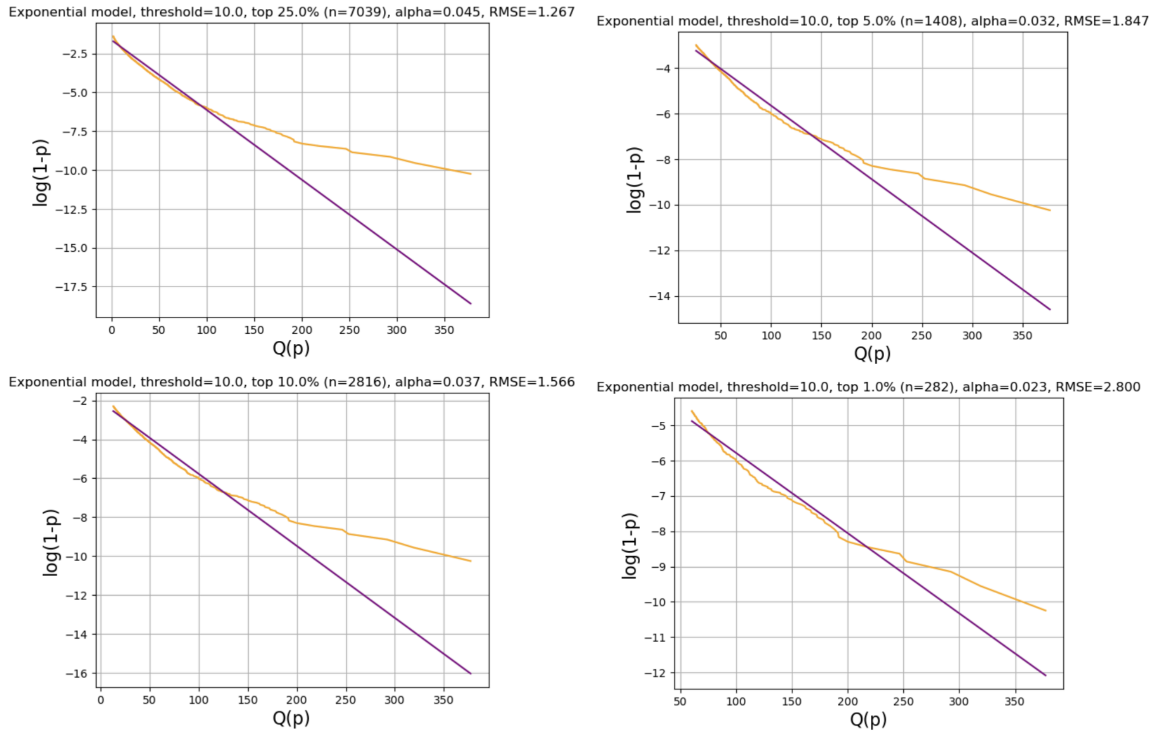

Powerlaw model, threshold=10.0, top 25.0% (n=7039), alpha=0.928, RMSE=1.426

Powerlaw model, threshold=10.0, top 5.0% (n=1408), alpha=2.271, RMSE=2.828

Powerlaw model, threshold=10.0, top 10.0% (n=2816), alpha=1.727, RMSE=1.851

Powerlaw model, threshold=10.0, top 1.0% (n=282), alpha=2.919, RMSE=4.189

Figure 1: Power-Law Fit Across Different Percentiles (10.0 mm Threshold).


Exponential model, threshold=10.0, top 25.0% (n=7039), alpha=0.045, RMSE=1.267

Exponential model, threshold=10.0, top 5.0% (n=1408), alpha=0.032, RMSE=1.847

Exponential model, threshold=10.0, top 10.0% (n=2816), alpha=0.037, RMSE=1.566

Exponential model, threshold=10.0, top 1.0% (n=282), alpha=0.023, RMSE=2.800

Figure 2: Exponential Fit Across Different Percentiles (10.0 mm Threshold).

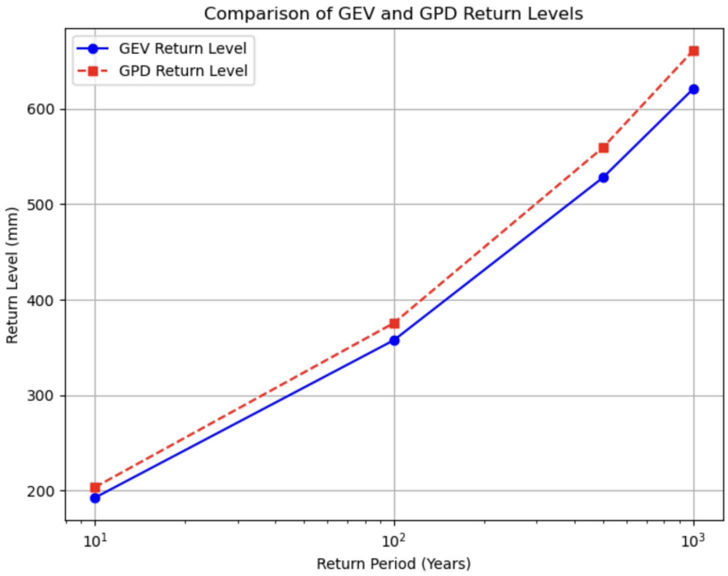


Figure 3: GEV and GPD Return Levels Comparison

Focus	
Methods	
Writing	
Findings	