Analyze Detroit's extreme snowfall risk using extreme value methods

1. Introduction

Extreme snowfall events, though rare, can be destructive to people's lives, especially in Detroit, where snowfall is more frequent than in other states. In this memo, we used extreme value methods to investigate whether Detroit's daily snowfall distribution exhibits heavy-tailed behavior and quantify the risk of extreme snowfall events. The analysis focuses on characterizing the statistical properties of the upper tail of the snowfall distribution, since the tail behavior indicates the long-term propensity to experience extreme weather events. By applying probability tail plots, L-moments, the Hill plot, and block maxima methods with Generalized extreme value(GEV) distributions, we assessed whether a heavy-tailed distribution best describes extreme snowfall in Detroit and estimated return levels that quantify the risk of events over 10 to 500 years.

2. Data Description

The dataset comprises 24,360 daily snowfall records in Detroit, spanning multiple decades. For each day, the record includes the calendar date, the total snowfall accumulated over 24 hours (measured in millimeters), and the corresponding year, which is used to define annual blocks for block-maxima analysis. A threshold equal to 100 mm is created for exceedance analysis.

3. Methods

To diagnose the tail behavior of Detroit snowfall, we applied three statistical tools. First, the probability tail plot serves as a diagnostic tool to evaluate whether the extreme upper-order statistics are more consistent with a power-law behavior or with an exponential pattern. In a heavy-tailed distribution, the tail probabilities tend to have power-law behavior and shrink more slowly than exponential decay. In the probability tail plot, given the threshold defined previously, if the data closely align with a power-law line in the log-log scale, whereas an exponential fit fails to capture the heavy tail, this indicates that the snowfall distribution is consistent with the heavy-tailed distribution. Second, L-Moment diagnostics capture the shape of the distribution. Specifically, the standardized fourth L-Moment is a measure of the tail heaviness. If the standardized fourth L-Moment exceeds 0.35, then the tails are sufficiently heavy. Third, the Hill estimator provides a nonparametric estimator of the tail index for a heavy-tailed distribution under the one-parameter Pereto assumption. By plotting the Hill estimate against the number of exceedances k, we can assess whether a stable flat region exists; the absence of such a region suggests that the data do not follow a global Pareto law.

To quantify extreme risk, the block maxima approach is applied to the data. Daily snowfall data is partitioned into annual blocks, and the maximum value in each block is kept. According to the Fisher-Tippett-Gnedenko theorem, the distribution of block maxima should be well approximated by the generalized extreme value distribution (GEV). A Q-Q plot shows the goodness of fit of the GEV to the data. Fitting the GEV to the block maxima allows the estimation of the return level, which represents the snowfall depth expected to be exceeded once every m years. These return levels provide interpretable risk measures for extreme snowfall in Detroit.

4. Results

Probability tail plots show the result of the exponential and power-law patterns for the top 0.5% of Detroit snowfall data. In the exponential pattern (Figure 1), the empirical tail diverges downward from the fitted straight line, indicating that the exponential distribution decays too rapidly and fails to capture the extreme snowfall events. In contrast, under the power-law pattern (Figure 2), the fitted line closely follows the empirical tail, producing an almost linear relationship in the log-log scale. The standardized fourth L-moment of Detroit snowfall data is 0.808. This value is higher than the conventional threshold of 0.35 that is used to indicate heavy-tailed behavior. The Hill plot (Figure 3) shows that the estimated tail index continuously decreases as the number of order statistics increases, without forming a stable flat region.

Using the block maxima approach, the GEV distribution fits the annual maxima well. The Q-Q plot (Figure 4) demonstrates a reasonably good fit between the empirical maxima and fitted quantiles for snowfall below approximately 300 mm, suggesting that the GEV captures the distribution of the moderate annual extremes (under 300 mm). However, for values exceeding 300 mm, the order statistics lie above the corresponding GEV quantile. The fitted GEV distribution yields return levels of 260.96mm (10 years), 302.47mm (20 years), 357.58mm (50 years), 399.93mm (100 years), and 501.36mm (500 years). For comparison, the exponential model produces slightly smaller return levels of 256.47 (10 years), 291.62mm (20 years), 338.09mm (50 years), 373.24mm (100 years), and 454.86mm (500 years).

5. Interpretation

In the probability tail plot, the strong alignment with the power-law fit and the deviation from the exponential fit indicate that the snowfall distribution exhibits heavy-tailed behavior. The standardized fourth L-moment, which exceeds 0.35, also supports this heavy tail behavior. Nevertheless, the absence of a flat region in the hill plot suggests that the snowfall distribution does not follow a global power-law distribution.

The block maxima analysis with the fitted GEV distribution further quantifies the risk through return levels. Specifically, daily snowfall exceeding 260mm is expected about once every 10 years, while more severe events exceeding 400mm are expected about once every 100 years. For comparison, the exponential model produces lower return levels. This divergence indicates that the exponential model underestimates the magnitude of extreme snowfall events relative to the GEV model, consistent with the evidence of heavy-tailed behavior from the tail diagnostic.

The Q-Q plot comparison suggests that the GEV distribution provides a good fit for most of the data range, but tends to underestimate the most extreme order statistics above 300mm. The observed extreme snowfall events in Detroit are more severe than those predicted by the fitted

GEV model. As a result, return level estimates for very long recurrence intervals, such as 500 years, should be interpreted with caution, as they may be biased downward.

6. Findings

In summary, our findings indicate that the analysis indicates that Detroit's daily snowfall distribution is best characterized as heavy-tailed. The standardized fourth L-moment and the probability tail plot both provide strong evidence of this conclusion. By contrast, the hill plot shows that the distribution is not a global Pareto distribution. The hill estimator is designed for one-parameter Pareto-type tails. When the true distribution is heavy-tailed but not strictly Pareto, the hill plot fails to show a flat region. The L-moment and probability tail plot are more general tools that capture heavy-tailedness even when the tail is not strictly Pareto. Therefore, we can still conclude that the snowfall distribution exhibits heavy-tailed behavior.

Moreover, the historical maximum daily snowfall is approximately 460mm, while the GEV return level analysis indicates that snowfall exceeding 400mm is expected about once every 100 years. Since our dataset only spans 1958-2025, longer observational records are needed to more rigorously validate the 460mm event against the estimated return levels. Nevertheless, the observed frequency of daily snowfall in the 200-300mm range aligns closely with the GEV-based return levels. This alignment strengthens confidence in the GEV model as a practical tool for extreme snowfall risk assessment.

Figure 1

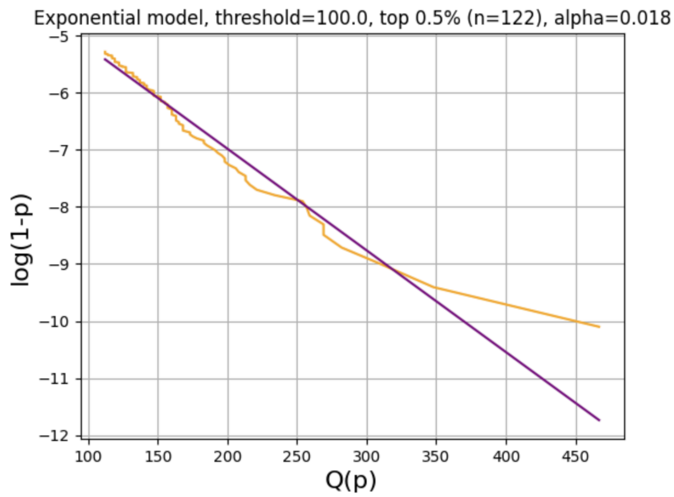


Figure 2

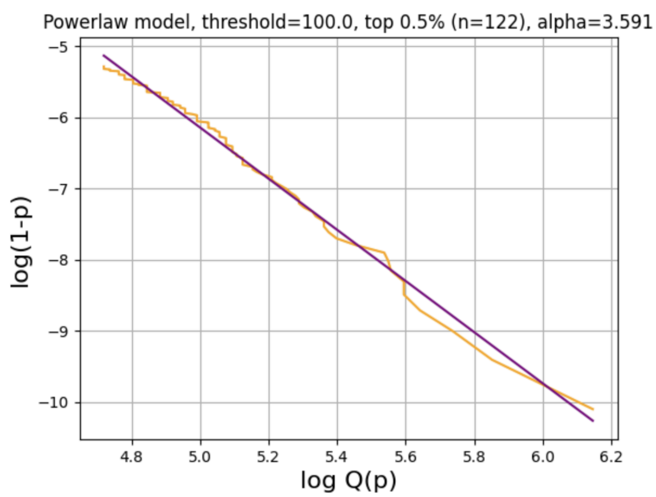


Figure 3

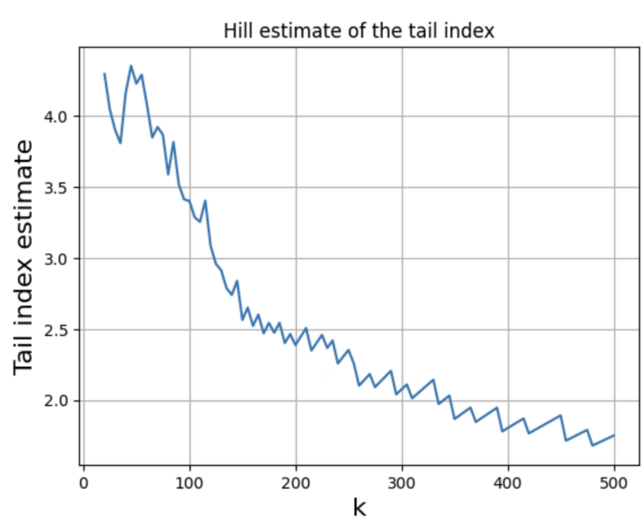
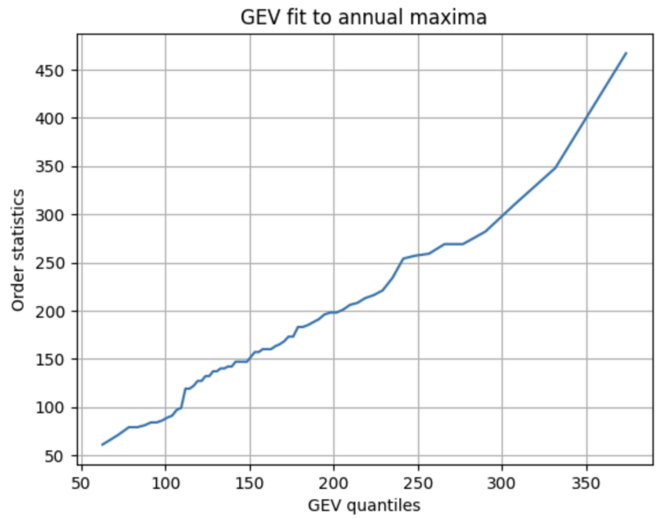


Figure 4


GEV fit to annual maxima
Order statistics
GEV quantiles
450
400
350
300
250
200
150
100
50
50
100
150
200
250
300
350