Extreme Rainfall Comparison in Lake and Sea Cities with Airports

Introduction

In this report, we undertake the task of constructing a generalized Pareto distribution model to analyze the exceedance of rainfall comparing cities adjacent to lakes versus those in proximity to the sea. Since both lake and sea cities are in proximity to water bodies, the question would arise: Would these two kinds of cities have the same extreme rainfall behavior? To be specific, would lake and sea cities have the similar tail indices or m-return? This problem is pivotal for shaping airport policies and strategies to effectively address the unique challenges posed by extreme rainfall events in diverse geographical contexts.

Data

For our analysis, we have harnessed meteorological data sourced from the National Centers for Environmental Information (NCEI) daily summaries dataset. To ensure a rigorous comparative evaluation, we selected five exemplary cities situated along the Great Lakes (Detroit, Madison, Chicago, Milwaukee, and Lansing) — the largest lake system in the United States. Furthermore, matching latitudes with their Great Lakes counterparts, we identified a set of five coastal cities (Baltimore, Philadelphia, New York, Hartford, and Boston) airports. This selection process enables us to investigate potential differences in tail indices while mitigating the influence of latitude-related variables. Our study encompasses data spanning from 1958 to 2023, as this time frame remains consistently available across all selected cities after the requisite data cleansing process.

Methodology

1.Pareto Distributions

As the simplest family of distributions with power law tails, Pareto distribution is common to estimate tail index. However, with only one parameter and the sample space of [1,inf), the Pareto distribution is always not adequate to come out with a good estimation.

2. Generalized Pareto Distribution

Generalized Pareto developed with two parameters would be beneficial to fit data well and sample space of [0,inf) also aligns reality. Pickands–Balkema–De Haan theorem also gives us confidence that with appropriate threshold, the exceedances for rainfall distributions approximately follow a generalized Pareto distribution. However, this method might challenge choosing the threshold and raise bias under positive serial dependence of data.

3. Generalized Extreme Value Distribution

One advantage of working with block-wise maxima and building GEV is that they are less sensitive to positive serial dependence that causes clusters of extreme values to occur in close proximity to each other. However, getting the block-wise maxima would shrink the valid number of data constraining our ability to accurately estimate the tail index.

Results

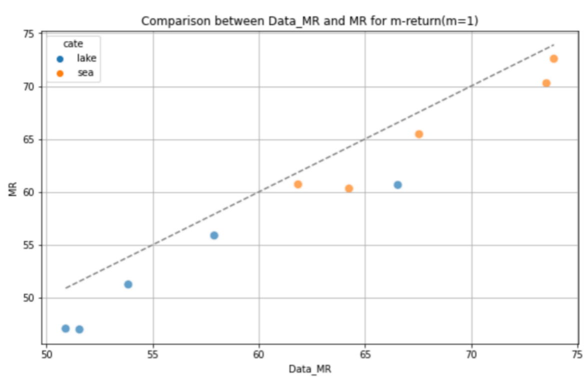
1. Model Selection

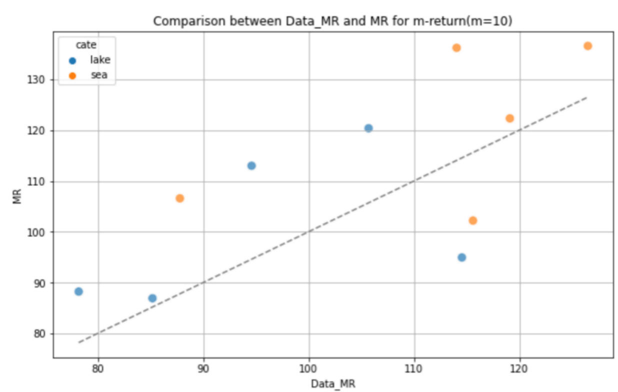
The Generalized Pareto Distribution to fit exceedances emerges as the best fitting choice for our analysis. By employing a threshold of 50mm, we gain access to a substantial dataset exceeding 20,000 observations. This larger dataset, coupled with the two parameters, affords a more adept characterization of the underlying data's behavior. The M return of single-parameter Pareto does not align with real data, which illustrates the limitation of Pareto Distribution. Furthermore, when we employ the block maxima approach and construct a Generalized Extreme Value Distribution, the m return deviates from the real data dramatically. This might be caused by limited availability of valid data, amounting to merely 65 records, making the bias tail index.

2. Fit of Generalized Pareto Distribution

Following the application of the Generalized Pareto Distribution (GPD) to the dataset encompassing ten airport cities, a critical assessment through QQ plots reveals a notably congruent alignment with an 45-degree straight line. This alignment underscores the appropriateness of the GPD model in characterizing the data.

Graph 1. Comparison of Data-based and Model-based M return

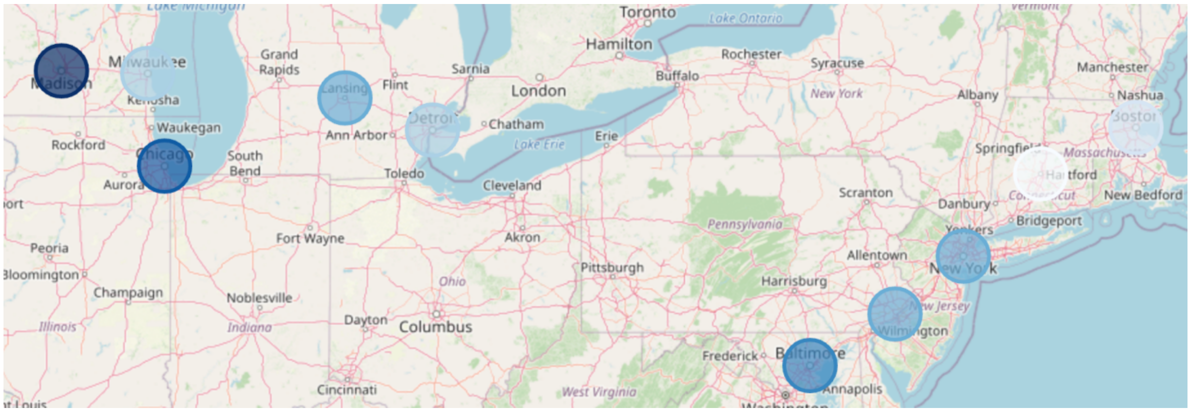




Graph 1 presents the m-return, calculated using the GPD, alongside real data for both 1-year and 10-year durations. Remarkably, for m=1 year, the data points illustrate bias with all Model-based M returns less than Data-based ones. But since Data-based m returns are not the exact definition of M returns and other methods have more severe bias, the GPD would still be the most promising one. When m=10 years, there is a discernible linear trend but with substantial scatter. However, considering the limitation of 10-year-return calculations on a relatively small dataset with 6-7 observations, such substantial deviations are expected.

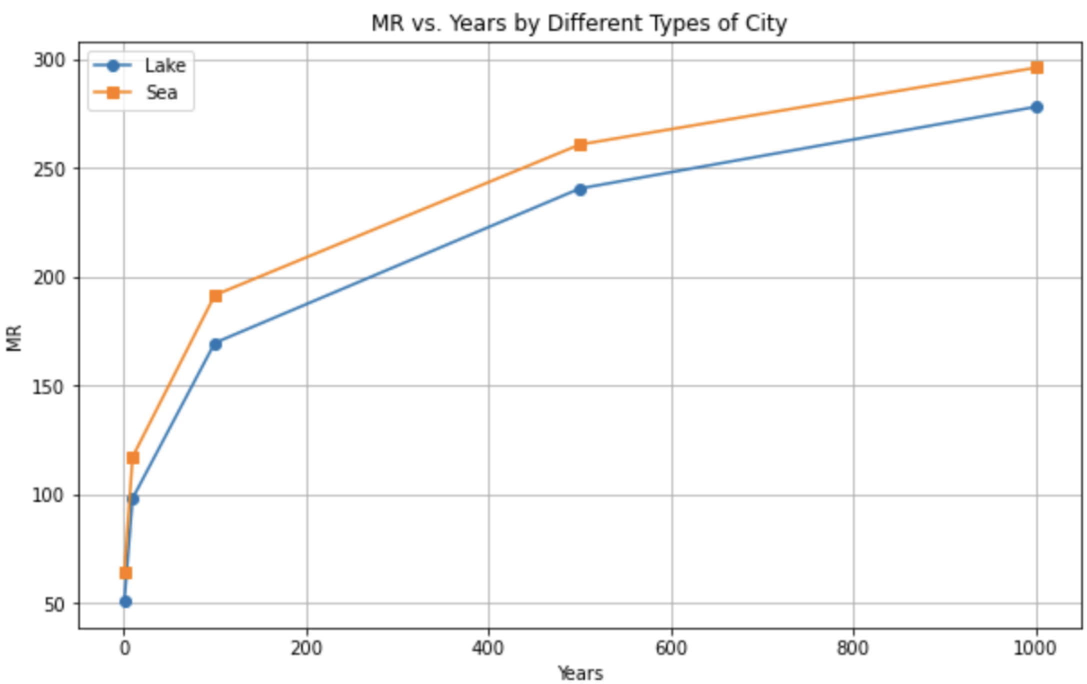
3. Findings and conclusions

Graph 2. Tail Index Map of Lake and Sea Cities



Graph 2 provides a visual representation of the tail index estimates derived from the Generalized Pareto Distribution for each city under consideration. The color gradient, with darker hues indicating larger tail indices. Notably, the tail indices for lake-adjacent cities consistently surpass those of their sea-adjacent counterparts. This observation implies that, within the analogous latitude range spanning from 1958 to 2023, the likelihood of encountering extreme rainfall events is notably higher in cities situated near the sea as opposed to those neighboring lakes.

Graph 3. Averaged M return for Lake and Sea Cities



Graph 3 elucidates the comparative analysis of average m-returns between sea and lake cities. Notably, the m-returns for sea cities consistently exceed those of their lake city counterparts across all observation periods. This substantiates a higher expectation value of extreme rainfall events occurring in sea cities over various temporal scales (1/10/100/1000 years).