# Risk Analytics - Practical 2

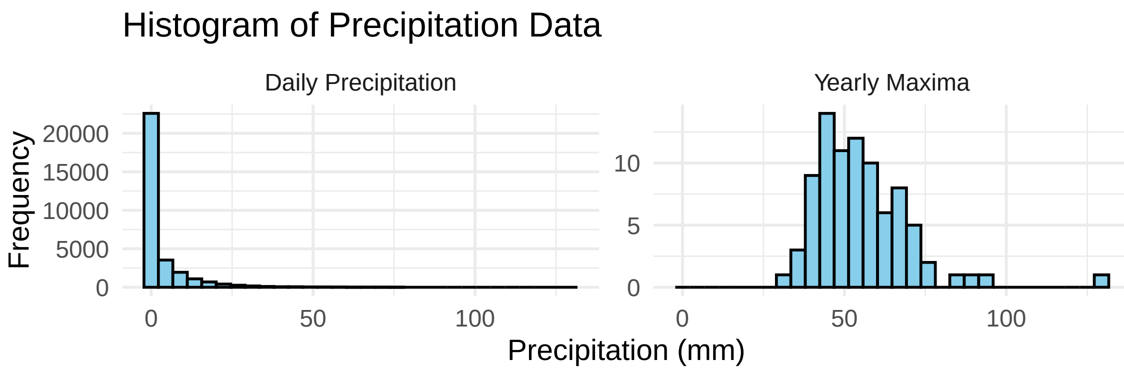
Winter semester 2024-2025, HEC, UNIL

Robin Michel, Faber Bickerstaffe, Antoine Magnin, Anastasia Pushkarev and Victorien Rodondi

2024-12-05

## Part 1: Block maxima approach

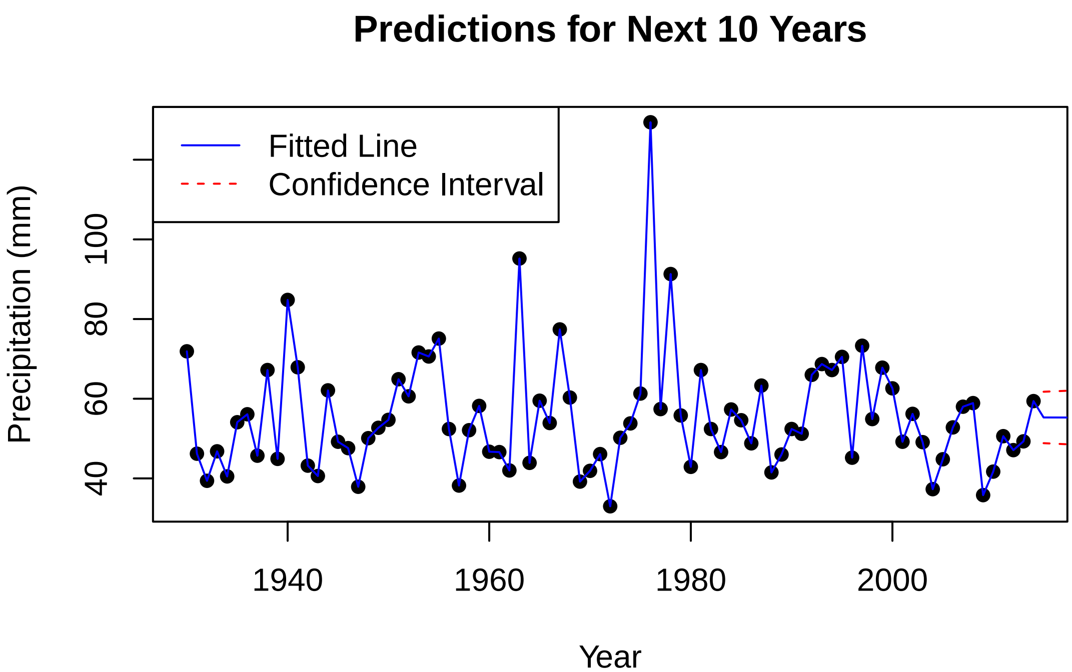
#### a) Read in the data and plot daily precipitation historgram & b) Extract yearly maxima and plot histogram




The majority of daily precipitation values are below 10 mm. Extreme precipitation values above 40 mm are rare but present. A Generalized Extreme Value (GEV) distribution may be suitable for the extremes, while a Gamma distribution better fits overall data.

The yearly maxima are right-skewed, with extreme values reaching above 120 mm. This suggests GEV modeling is appropriate for analyzing these extremes.

**c) Fit a linear model to yearly maxima and predict next 10 years**




The linear model suggests a steady increase in yearly maximum precipitation. This method sees oversimplify the complexities of extreme precipitation patterns.

**d) Fit GEV models and compare AIC/BIC**

```
## AIC (Constant Parameters): 672.9433
```

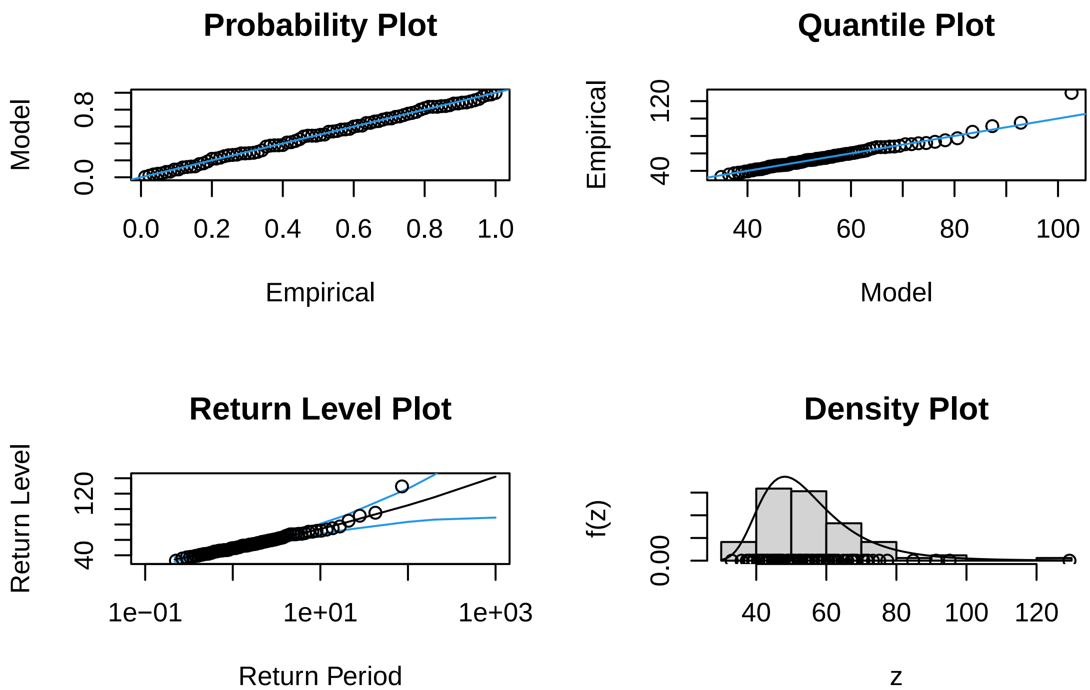
```
## AIC (Time-Varying Location): 674.8906
```

```
## BIC (Constant Parameters): 680.2712
```

```
## BIC (Time-Varying Location): 684.6612
```

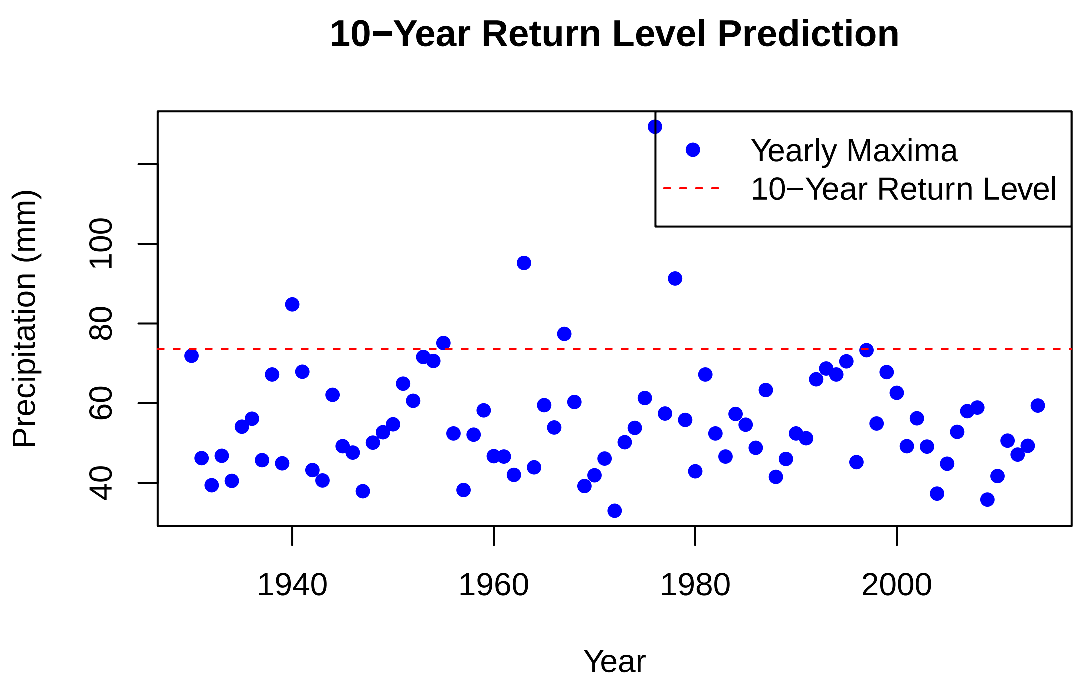
The constant GEV model has slightly lower AIC and BIC values, indicating better fit compared to the time-varying model. Therefore, the constant model is recommended.

**e) Diagnostic plots of GEV fit**




Diagnostic plots suggest the model fits the data well, as evidenced by the quantile and return-level plots. Slight deviations at extremes should be noted, as they may affect predictions.

**f) Predict the 10-year return level and plot**




The 10-year return level is approximately 73.61 mm. Few historical events exceed this level.

**g) Count exceedances for return levels**

```
##        10        20        50        85
##  73.60759  82.53069  94.90440 102.45270
```

```
## 10 20 50 85
##  6  4  2  1
```

The historical counts above the 10-, 20-, 50-, and 85-year return levels are 6, 4, 2, and 1 respectively.

**h) Return period for 100 mm of precipitation**

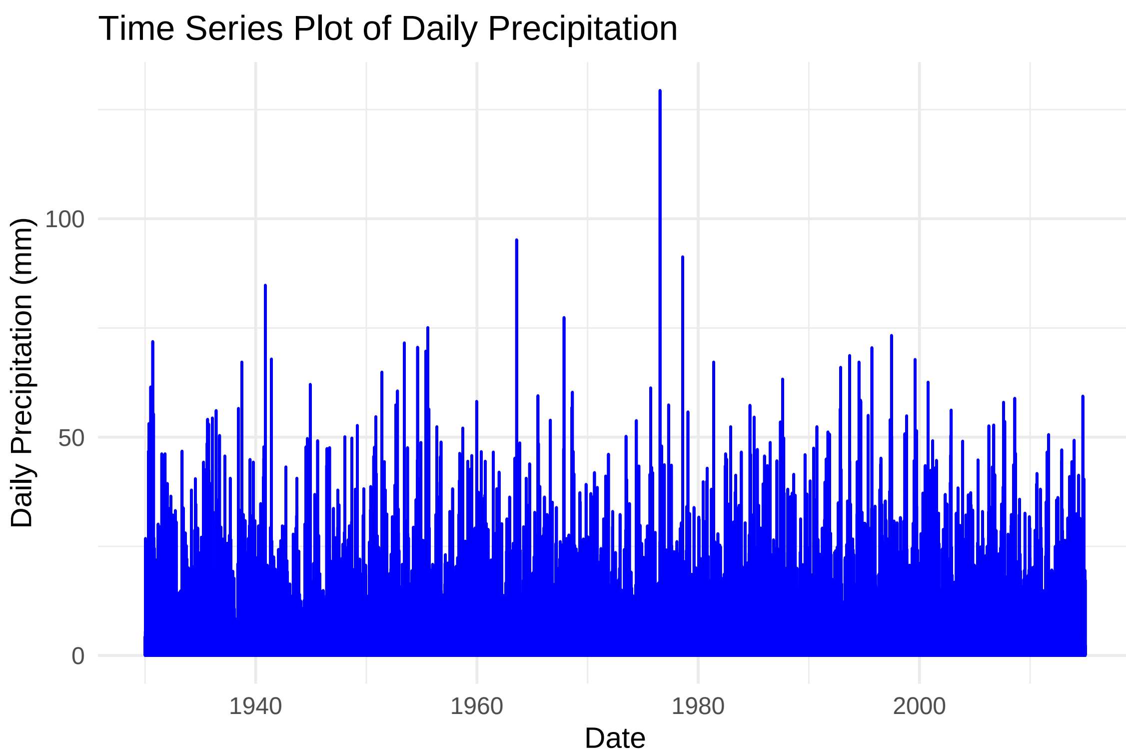
```
## Return period for 100 mm precipitation: 71.70624 years
```

**i) Probability of exceeding 150 mm in a given year**

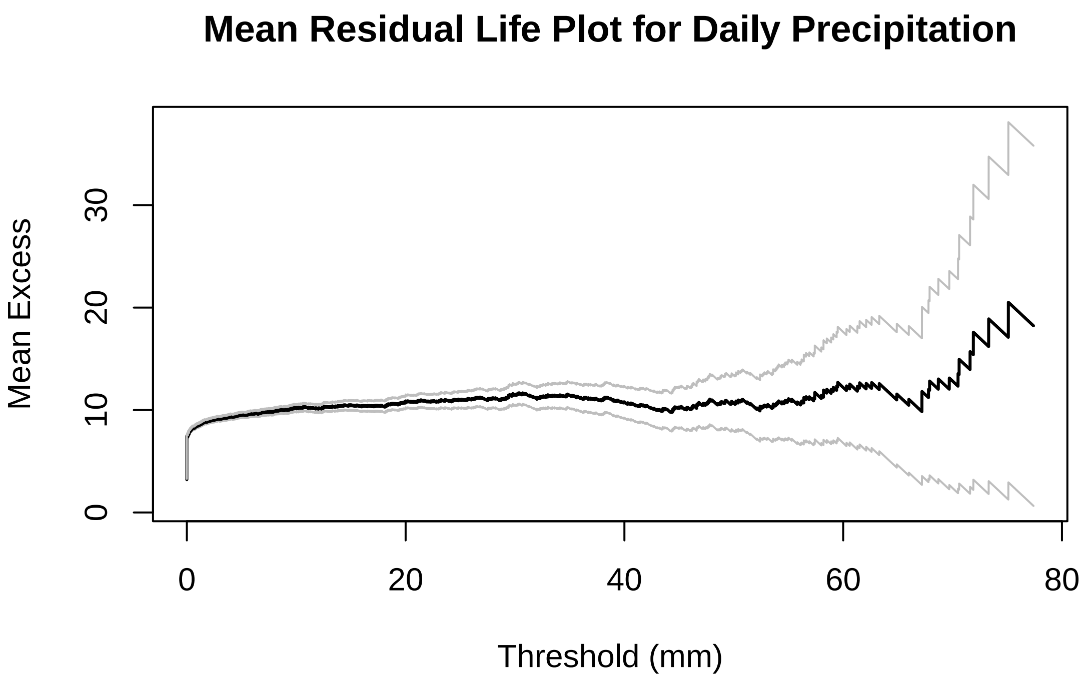
```
## Probability of exceeding 150 mm in a day at least once in a year: 0.2094091
```

## Part 2: Peaks-Over-Threshold Approach

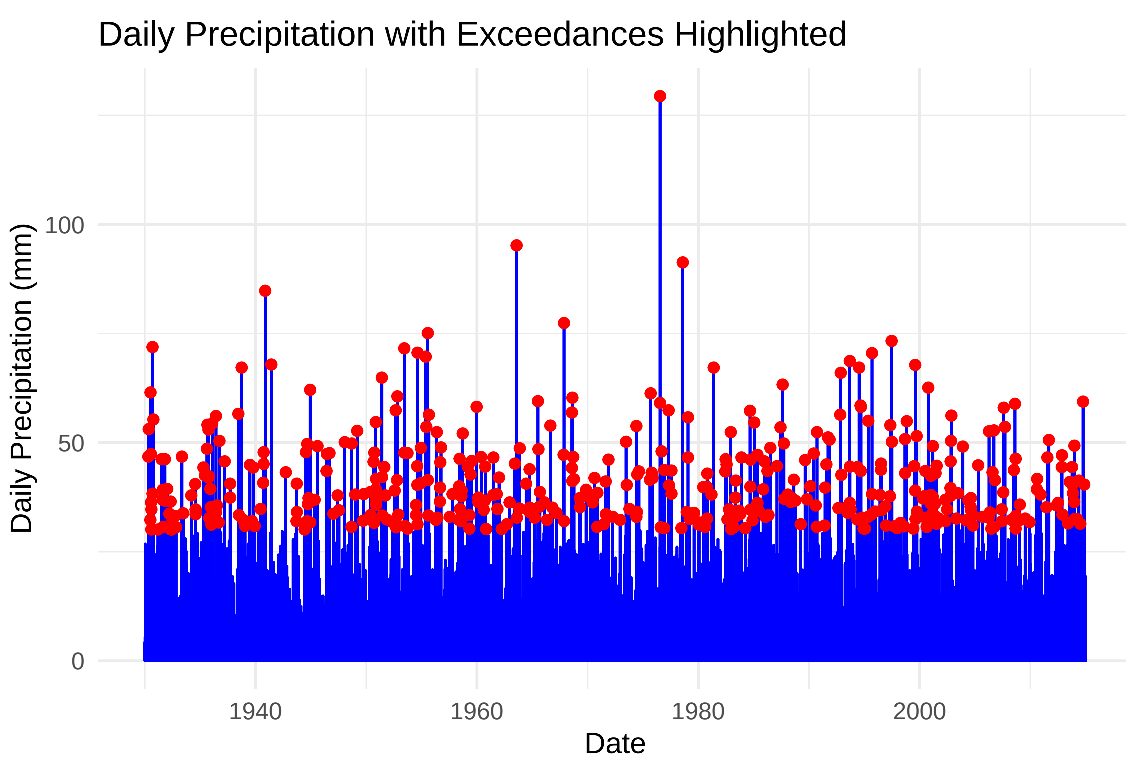
**a) Time series plot of daily precipitation**

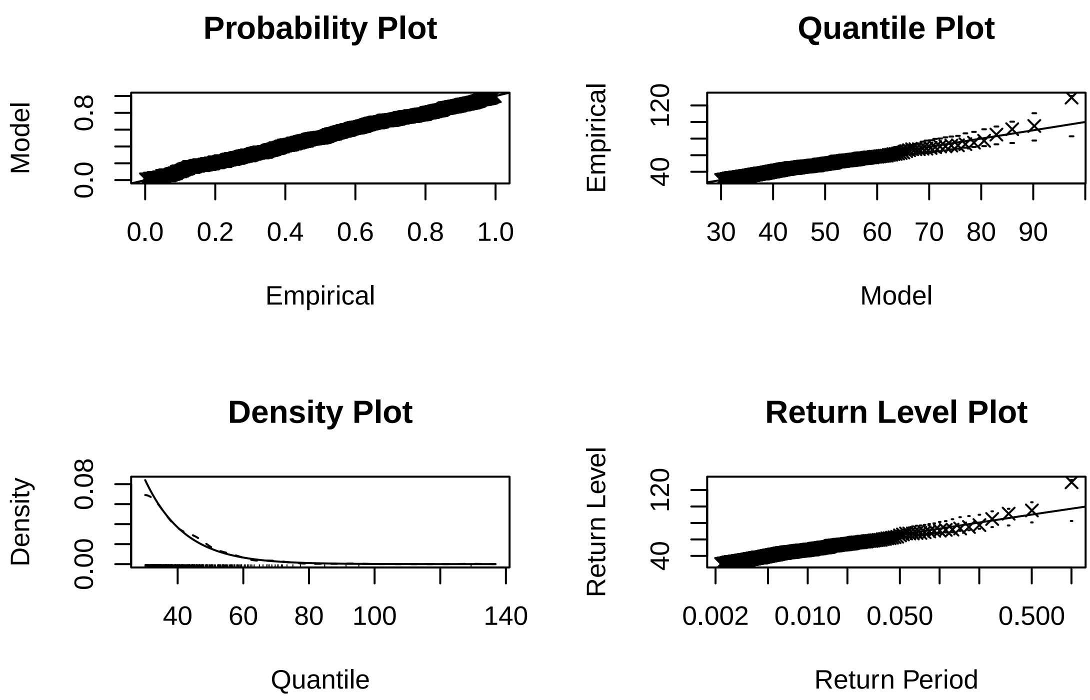


**b) Mean Residual Life Plot and Threshold Selection**




**c) Fit a Generalized Pareto Distribution (GPD) and the data exceeding the threshold**






**d) Return Levels for Different Periods**

```
## Return levels for specified return periods (in mm):

##       10       20       50       85
## 101.5261 108.5137 117.5826 122.7484
```

**e) Return Period for 100 mm Precipitation**

```
## Return period for 100 mm precipitation: 68.38326 years
```

**f) Probability of Exceeding 150 mm in a Given Year**

```
## Probability of exceeding 150 mm at least once in a year: 0.003411695
```

**g) Comparison of POT and Block Maxima Methods**

## Comparison of POT and Block Maxima Methods

### Advantages of POT Approach

- More data points: Uses all exceedances over a threshold, providing more data for analysis and improving parameter estimation.
- Better tail modeling: Focuses on extreme data, making it more effective for modeling the tail of the distribution, which is critical for extreme event analysis.

### Drawbacks of POT Approach

- Threshold selection: Requires careful selection of a threshold, which can be subjective and significantly impact model fit.
- Sensitivity: Results can be highly sensitive to the chosen threshold, potentially leading to biased estimates if the threshold is poorly chosen.

#### Advantages of Block Maxima Method

- Simplicity: Conceptually simple and widely understood, involving the selection of maximum values from defined blocks (e.g., annual maxima).
- Practical focus: Often focuses on annual maxima, which can be of direct practical interest for many risk assessment applications.

#### Drawbacks of Block Maxima Method

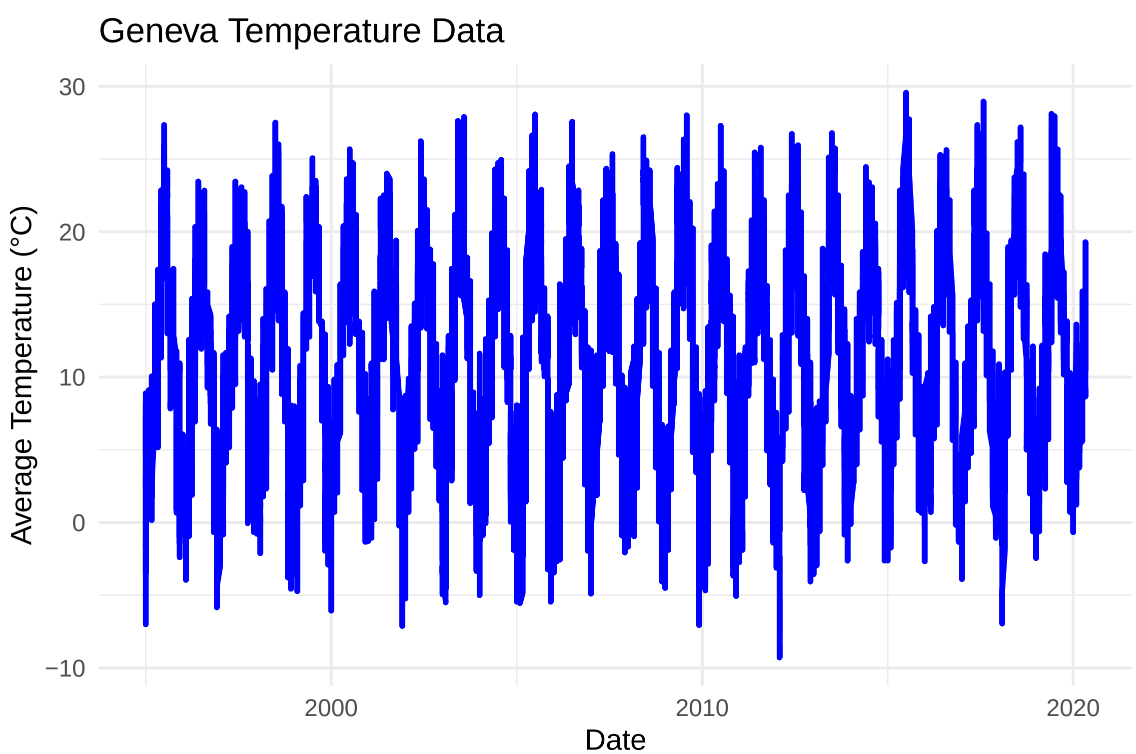
- Data inefficiency: Discards all but the maximum value from each block, leading to a loss of information, especially when more extreme values exist within the block.
- Higher variance: Due to fewer data points, the resulting estimates tend to have larger variance compared to the POT approach.

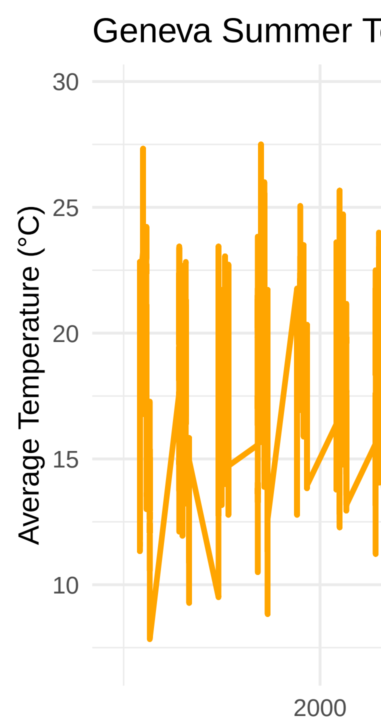
#### Preference

The POT method is generally preferred when the objective is to make full use of available extreme data and a good threshold can be selected. However, the Block Maxima method is simpler and often sufficient for practical purposes, especially when clear block segmentation exists (e.g., annual maxima).

## Part 3: Clustering and Seasonal Variations

#### a) Upload the Geneva temperature data. Plot the data. Subset the data for the summer months (June to September).




This first graph shows the average monthly temperature in Geneva over the years. The temperatures highlight seasonal fluctuations and year-to-year variations.

The second graph shows the average monthly temperature in Geneva from June to September over the years. We can see the trend in the fluctuation for each month.

**b) Compute the extremal index of the subsetted series with appropriatelly chosen threshold. Do the extremes occur in clusters? What is the probability that if the temperature today is extreme (above the chosen threshold) then tomorrow will be also extreme?**

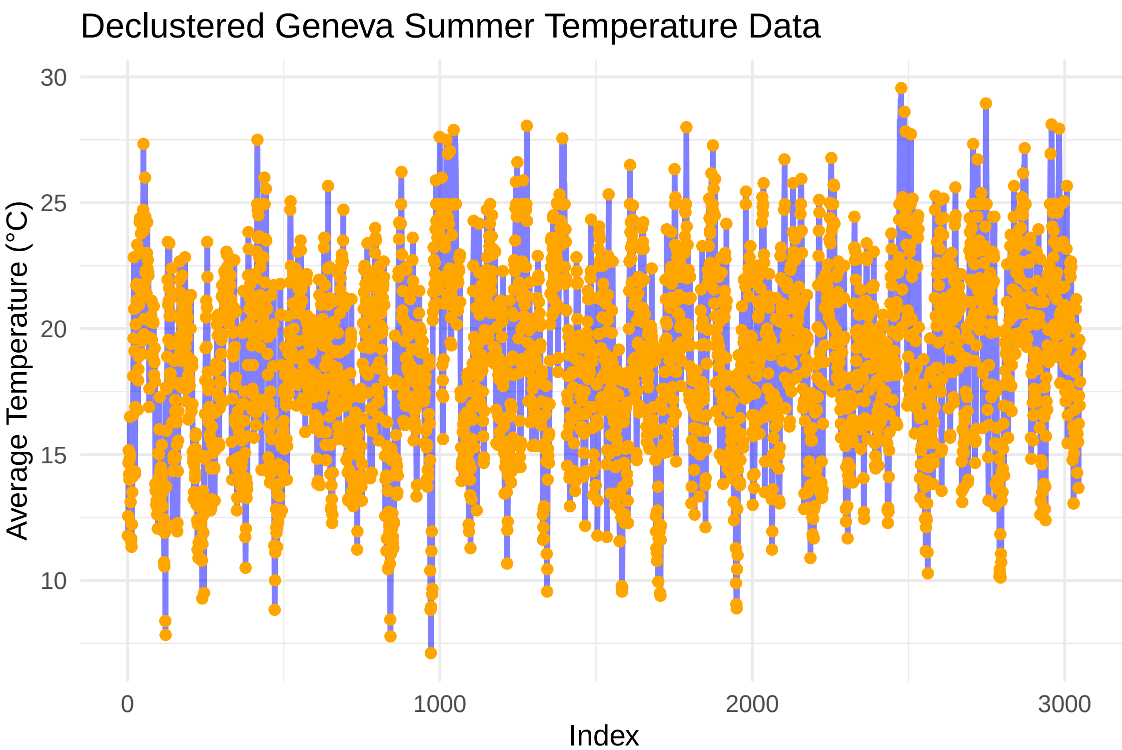
```
## Extremal Index: 0.2612517 40 9
```

```
## Do extremes occur in clusters?  Yes No No
```

```
## Probability that if today's temperature is extreme, tomorrow's will be also extreme: 0.2612517 40 9
```

The threshold for extreme temperatures is set at the 95th percentile of the data, and the extremal index is computed, which quantifies how extremes are distributed—values below 1 suggest clustering, meaning extremes are likely to occur in groups. Additionally, the extremal index is interpreted as the probability that if today's temperature is extreme, tomorrow's will also be extreme, with a value of 0.2612517 in this case.

**c) Decluster the data using a suitable threshold. Plot the resulting declustered data.**




We applied a threshold at the 95th percentile to identify independent extreme events and removed clusters of extremes. The resulting declustered data is plotted alongside the original data, showing how extreme temperatures are isolated after declustering.

The graph displays the original Geneva summer temperature series as a blue line, with declustered extreme temperatures highlighted as orange dots. The declustered points represent independent extreme events that surpass the threshold.

**d) Fit a Generalized Pareto Distribution (GPD) to the data, both raw and declustered. Compare the models and compute 10-year return level.**

```
##
## fevd(x = summer_data$AvgTemperature, threshold = threshold, type = "GP",
##     method = "MLE")
##
## [1] "Estimation Method used: MLE"
##
```

```
##
##  Negative Log-Likelihood Value:  183.7343
##
##
##  Estimated parameters:
##      scale      shape
##  1.6646439 -0.3008329
##
##  Standard Error Estimates:
##      scale      shape
## 0.17468588 0.07091298
##
##  Estimated parameter covariance matrix.
##             scale        shape
## scale  0.03051516 -0.010674213
## shape -0.01067421  0.005028651
##
##  AIC = 371.4686
##
##  BIC = 377.5164
```

```
##
## fevd(x = declustered_data, threshold = threshold, type = "GP",
##     method = "MLE")
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  85.94948
##
##
##  Estimated parameters:
##      scale      shape
##  1.8667514 -0.3413717
##
##  Standard Error Estimates:
##     scale     shape
## 0.2932640 0.1072101
##
##  Estimated parameter covariance matrix.
##             scale       shape
## scale  0.08600374 -0.02772025
## shape -0.02772025  0.01149401
##
##  AIC = 175.899
##
##  BIC = 180.3083
```

```
## 10-year Return Level (Raw Data): 29.32159
```

```
## 10-year Return Level (Declustered Data): 29.18888
```

The results show that the raw data model has slightly higher scale and shape parameter estimates compared to the declustered model, leading to a 10-year return level of 29.32°C for the raw data and 29.19°C for the declustered data. The declustering reduces clustering bias in extremes, resulting in a slightly more conservative estimate of extreme temperature levels. The AIC and BIC values also confirm better fit for the declustered model, indicating improved reliability for return level predictions.