Risk Analytics
Practical 2

Frédéric Aviolat and Juraj Bodik

Winter semester 2025-2026, HEC, UNIL

River Discharge in Lake Neuchâtel region

Figure 1: **Flood in Mesolcina valley** (Graubünden) on 23.6.2024

Lake Neuchâtel region faces significant hydrological challenges due to fluctuations in river discharge, particularly in response to seasonal weather patterns and extreme rainfall events. Sudden spikes in river discharge can lead to flooding, erosion, and disruption of infrastructure. It is estimated that river discharge above 100 m^3/s in a single day may critically stress the region's drainage and flood control systems.

In this practical, we will analyze daily river discharge data in Lake Neuchâtel region from 1930 to 2014 and estimate the probability of a day exceeding a critical threshold of 100 m^3/s.

Part 1: GEV approach

Your goal is to model and analyze the discharge extremes using the block maxima approach.

(a) Extract the yearly maximum discharge values. Draw their histogram. Which well-known distribution does it resemble? Compare with your answer from Practical 1, Part 1(b).

(b) Fit a basic linear model to the yearly maximum discharge values ($discharge \sim time$) and predict the values for the next 10 years. Provide prediction intervals for your predictions and plot them. Do you think this is a reasonable approach?

(c) Fit a GEV model with constant parameters to the historical yearly maxima. We recommend using the `fevd` function from the `extRemes` package or the `gev.fit` function from the `ismev` package. Then fit a second GEV model with a time-varying location parameter. Compare the two models using AIC or BIC. Which one do you recommend?

(d) Draw diagnostic plots of your GEV fit (e.g., using `gev.diag`). Is it a good fit?

(e) Using your preferred model, predict the 10-year return level and calculate confidence bands for this prediction. Plot the return level estimates with point-wise confidence intervals together with the historical data.

(f) Broadly speaking, in any given year, there is a 1-in-10 chance of exceeding the 10-year return level. How many historical years exceed the 10-year return level? Repeat for the 20-year, 50-year, and 85-year return levels.

(g) Using the fitted model, compute the probability that the river discharge exceeds 100 m^3/s on at least one day in the next year.

Part 2: Peaks-over-threshold approach

In this part, we use the Peaks-Over-Threshold (POT) approach to analyze extreme discharge events in Lake Neuchâtel region.

(a) Display a time series plot of the daily river discharge across the full data range.

(b) To model high discharge levels using the POT approach, first choose a threshold. Draw a Mean Residual Life Plot (e.g., using `mrlplot` from the `POT` package). Choose a reasonable threshold and highlight threshold exceedances in the time series plot from (a).

(c) Fit a GPD to the threshold exceedances and draw diagnostic plots. Is it a reasonable fit? (Hint: if not, reconsider the choice of threshold.)

(d) Using the fitted model, compute the 10-year, 20-year, 50-year, and 85-year return levels.

(e) Using the fitted model, compute the probability that the discharge exceeds 100 m^3/s on at least one day in the next year.

(f) Compare the results with the block maxima (GEV) method. Discuss the advantages and limitations of using the POT method. Which method do you prefer and why?

Part 3: Clustering and Seasonal Variations

In this part, we analyze whether extreme river discharge events occur in clusters and how they vary by season.

(a) Compute the extremal index of the seasonal subset using an appropriately chosen threshold (e.g., with the `extremalindex` function in the `extRemes` package). Do extreme values tend to cluster? What is the probability that if today's discharge is extreme, tomorrow's will be as well?

(b) Decluster the data using the chosen threshold (e.g., with the `decluster` function from `extRemes`). Plot the resulting declustered data.

(c) Fit a Generalized Pareto Distribution (GPD) to both the raw and declustered data. Compare the two models and compute the 10-year return level. How does declustering affect the estimated tail behavior?