ASSIGNMENT 2 - STATISTICAL METHODS

LGCIV 2055
Analysis and mitigation of floods

Professor : Sandra Soares Frazao
Teaching assistants : Nathan Delpierre, Martin Petitjean

Group H
Foret Félix
Herphelin Louis

Year 2025 - 2026

1 Introduction

This report presents a statistical analysis of the annual maximum discharges of the Ourthe River at the Hotton station (Belgium). The objective is to estimate characteristic flood discharges for different return periods and to assess the influence of the 2021 flood event on these estimates. The data used come from the Walloon Region Hydrology Office and cover the period from January 1, 1979 to September 16, 2025. The analysis was performed with Python.

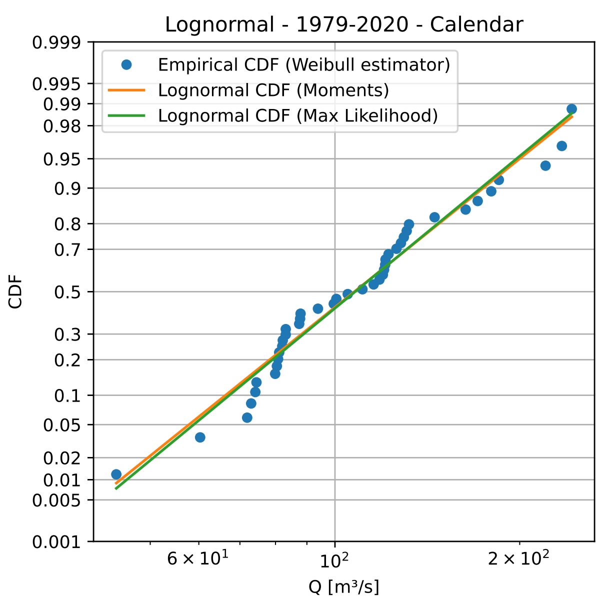
Our python code, as well as the output files (plots, Excel spreadsheet) are available on
`https://github.com/LLouuiss/Floods`

2 Analysis over the years 1979–2020

In this section, we perform a statistical analysis of annual maximum discharges using three probability distributions: the Lognormal distribution (Method of Moments & Maximum Likelihood) and the Gumbel distribution. For this study, two different types of data segmentation were considered: Calendar years (from January to December of each year) and hydrological years (from September to August).

The parameters of the lognormal distribution (mean m and variance σ^2 can be estimated using two classical approaches: the method of moments ($\hat{m}_{\ln Y} = \ln(m_Y) - \frac{1}{2}\sigma^2_{\ln Y}$ and $\hat{\sigma}^2_{\ln Y} = \ln\left(\frac{\sigma_Y^2}{m_Y^2} + 1\right)$) and the estimation of maximum likelihood ($\hat{m}_{\ln Y} = \frac{1}{n}\sum_{i=1}^{n} \ln(Y_i)$ and $\hat{\sigma}^2_{\ln Y} = \frac{1}{n}\sum_{i=1}^{n} \left(\ln(Y_i) - \hat{m}_{\ln Y}\right)^2$). The two parameter of the Gumbel distribution (α and u) can also be estimated with the mean and variance of the dataset ($u = m_Y - \frac{0.577}{\alpha}$ and $\alpha = \frac{\pi}{\sqrt{6}\sigma_Y}$). Once the parameters are estimated, the theoretical Cumulative Distribution Function (CDF) can be computed for each model and compared to the empirical data obtained using the Weibull estimator, $F(X_i) = \frac{i-0.5}{n}$.

Tables 2.1 and 2.2 summarize the parameters obtained for each distribution and year type. The Figures 2.1 and 2.2 illustrate the comparison between the empirical and theoretical distributions.



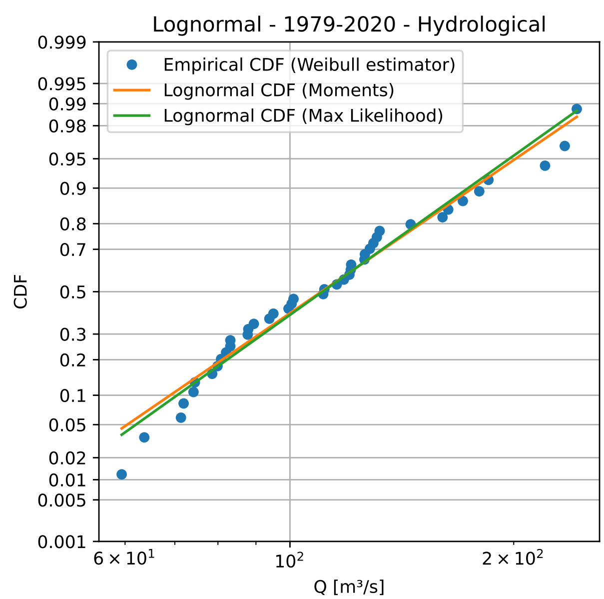


Figure 2.1: Comparison between observed and theoretical distributions
(Lognormal, Moments & Maximum Likelihood methods).

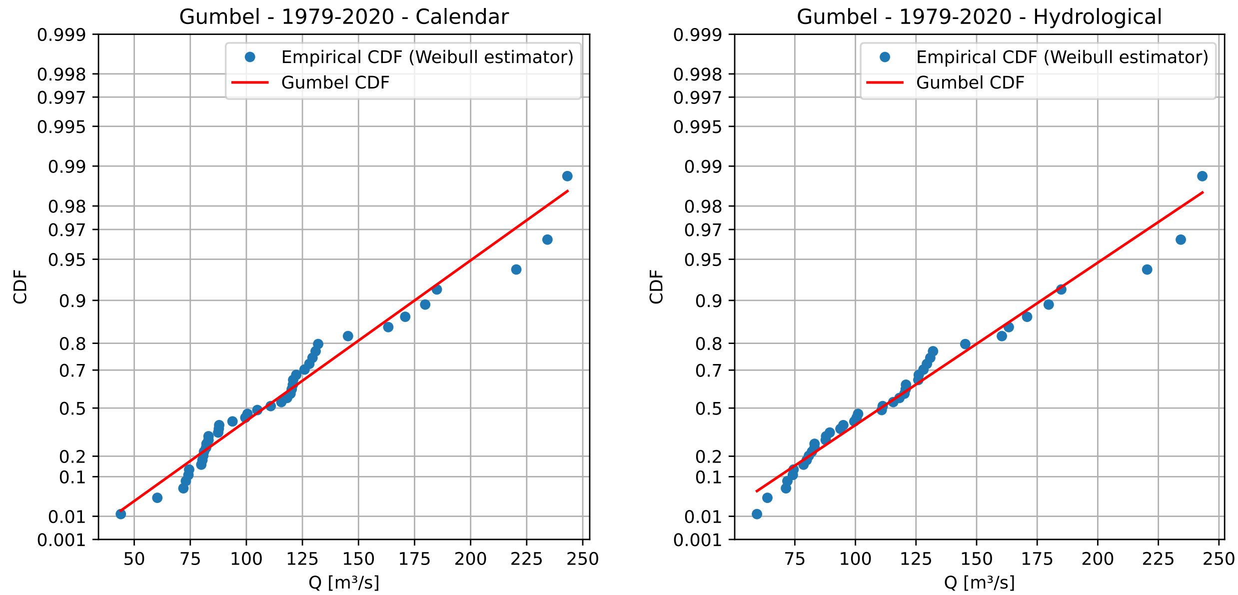


Figure 2.2: Comparison between observed and theoretical distributions (Gumbel).

As can be seen from the figures, our data seem to fit with all three distributions. We also note that the dataset separated by hydrological years fits better than when using calendar years. It is because hydrological years better capture the flood season avoiding splitting flood events that occur between December and January.

Our previous remark was graphical; therefore, to quantitatively evaluate the goodness of fit of these distributions to the dataset, we apply two tests: the Kolmogorov–Smirnov ($\alpha = 0.1$, the value recommended in our syllabus) test and the Chi-squared test ($\alpha = 0.05$, also recommended in our syllabus). A test is considered passed if the probability that the data follow the given distribution is $1-\alpha$. The Kolmogorov–Smirnov (K–S) test evaluates the maximum deviation between the theoretical cumulative distribution function (CDF) and the empirical one built from the observed data. The test statistic corresponds to the largest vertical distance between these two curves. If this deviation remains smaller than a critical value C_α (which depends on the confidence level $1-\alpha$ and the sample size n), the null hypothesis that the data follow the assumed distribution cannot be rejected. The Chi-squared (χ^2) test compares the number of observations in each predefined class with the number theoretically expected according to the tested distribution. The sum of the squared differences, normalized by the expected frequencies, gives the test statistic. If this value remains below the critical threshold $\chi^2_{\alpha,k-p-1}$ (where k is the number of classes and p the number of estimated parameters), the hypothesis of adequacy is accepted. This method provides a more global view of the discrepancies between observed and theoretical frequencies. For the Chi-squared test, we perform it twice: once using 4 classes and once using 5 classes. The results are presented in the table below. Together, these two complementary tests allow both a pointwise (K–S) and classwise (Chi-squared) verification of the goodness of fit of the selected distributions.

Finally, the extreme flows corresponding to different return periods are calculated and summarized in Tables 2.1 and 2.2.

	LN moments	**LN maxi likelihood**	**Gumbel moments**
Parameters	$\hat{m}_{\ln Y} = 4.678$	$\hat{m}_{\ln Y} = 4.681$	$u = 94.907$
	$\hat{\sigma}_{\ln Y} = 0.377$	$\hat{\sigma}_{\ln Y} = 0.368$	$\alpha = 0.0281$
Extreme discharges			
$T_r = 10$ years	174 m^3/s	173 m^3/s	175 m^3/s
$T_r = 100$ years	259 m^3/s	254 m^3/s	259 m^3/s
$T_r = 1000$ years	345 m^3/s	337 m^3/s	341 m^3/s
$T_r = 10000$ years	437 m^3/s	424 m^3/s	423 m^3/s
K-S test			
$\alpha = 0.10$	$0.108 < 0.188$	$0.116 < 0.188$	$0.109 < 0.188$
Chi squared			
4 classes ($\alpha = 0.05$)	$1.539 < 3.842$	$1.779 < 3.842$	$1.648 < 3.842$
5 classes ($\alpha = 0.05$)	$4.470 < 5.992$	$5.101 < 5.992$	$4.565 < 5.992$

Table 2.1: Calendar years

	LN moments	**LN maxi likelihood**	**Gumbel moments**
Parameters	$\hat{m}_{\ln Y} = 4.703$	$\hat{m}_{\ln Y} = 4.706$	$u = 97.561$
	$\hat{\sigma}_{\ln Y} = 0.366$	$\hat{\sigma}_{\ln Y} = 0.352$	$\alpha = 0.0284$
Extreme discharges			
$T_r = 10$ years	176 m^3/s	174 m^3/s	177 m^3/s
$T_r = 100$ years	258 m^3/s	251 m^3/s	259 m^3/s
$T_r = 1000$ years	341 m^3/s	328 m^3/s	383 m^3/s
$T_r = 10000$ years	430 m^3/s	409 m^3/s	478 m^3/s
K-S test			
$\alpha = 0.10$	$0.097 < 0.188$	$0.094 < 0.188$	$0.099 < 0.188$
Chi squared			
4 classes ($\alpha = 0.05$)	$1.675 < 3.842$	$2.034 < 3.842$	$1.763 < 3.842$
5 classes ($\alpha = 0.05$)	$7.254 \nless 5.992$	$7.957 \nless 5.992$	$7.389 \nless 5.992$

Table 2.2: Hydrological years

We observe that the estimated discharges logically increase with the return period. The Gumbel distribution generally provides the highest discharge values for a given return period when considering hydrological years, whereas this is not always the case for calendar years. The Lognormal (Method of Moments) distribution consistently gives the lowest discharges for hydrological years, which is not systematically true for calendar years. Regarding the goodness-of-fit tests, all tests are passed for the calendar years, while for the hydrological years the Chi-squared test with five classes fails, whereas the others are accepted. Overall, all tests tend to give the lowest statistic values for the Lognormal (Method of Moments) distribution, confirming its slightly better fit to the observed data.

3 Results for 1979–2025

To continue our analysis, we take into account the influence of the 2021 flood by extending our data set to include data up to 2025. To do this, we applied the same procedure as in the previous section to evaluate the characteristic discharges.

The cumulative distribution functions computed for the years 1979-2025 is represented in Figure 4.1. The point on the extreme right is the discharge corresponding to the 2021 flood event. We see that the theoretical CDF fits less accurately the empirical data.

The characteristic discharge y_r for a given return period T_r is the discharge such that the probability of not having a higher discharge during a period of T years is $\left(1 - \frac{1}{T_r}\right)^T$

$$P[(Y \leq y_r)_T] = \left(1 - \frac{1}{T_r}\right)^T$$

Alternatively, the return period for a given discharge y_r is the value T_r for which the equation above is true.

In this case, we considered a period T of 1 year.

We also included the return period of the maximum flood discharge observed in 2021 to highlight how exceptional this event was, with $Q_{\text{max-2021}} = 327$ [m^3/s]. All results are summarized in the Table 4.1. Here, only calendar years are considered.

Including the 2021 flood shifts the fitted distributions, increasing the estimated discharges for all return periods, and reducing the return period of the 2021 event, as shown by Figure 4.2. This behavior is typical for short datasets: A single exceptional flood can significantly bias the statistical estimation, highlighting the limitations of purely statistical approaches when we dispose of a small amount of samples (here, 47 discharges) and extreme events are rare.

4 Conclusion

We compared the statistical analyses of the annual maximum discharge in the river Ourthe at Hotton, one over the years 1979-2020 and the other over the years 1979-2025 (including the extreme flood that occurred in July 2021).

First, we confirmed the validity of using Lognormal and Gumbel distributions to model the repartition of discharges, through Kolmogorov-Smirnov and Chi-squared test that passed excepted for 5-class test on hydrological years.

We then observed that, in the second case, the estimated characteristics floods with return periods of 10, 100, 1000 & 10000 years were significantly higher than in the first case. Consequently, the return period for the discharge of July 2021 (327 m^3/s) was more than halved when taking into account the data of 2021-2025.

Period	**Distribution**	Q_{10} $[m^3/s]$	Q_{100} $[m^3/s]$	Q_{1000} $[m^3/s]$	Q_{10000} $[m^3/s]$	**Return period (2021)** T_r [years]
1979–2020	LN (mom)	174	259	345	437	630
	LN (ML)	173	254	337	424	775
	Gumbel	175	259	341	423	679
1979–2025	LN (mom)	187	292	403	526	213
	LN (ML)	181	270	363	463	417
	Gumbel	189	287	383	478	264

Table 4.1: Comparison of extreme discharges with and without 2021 flood event.

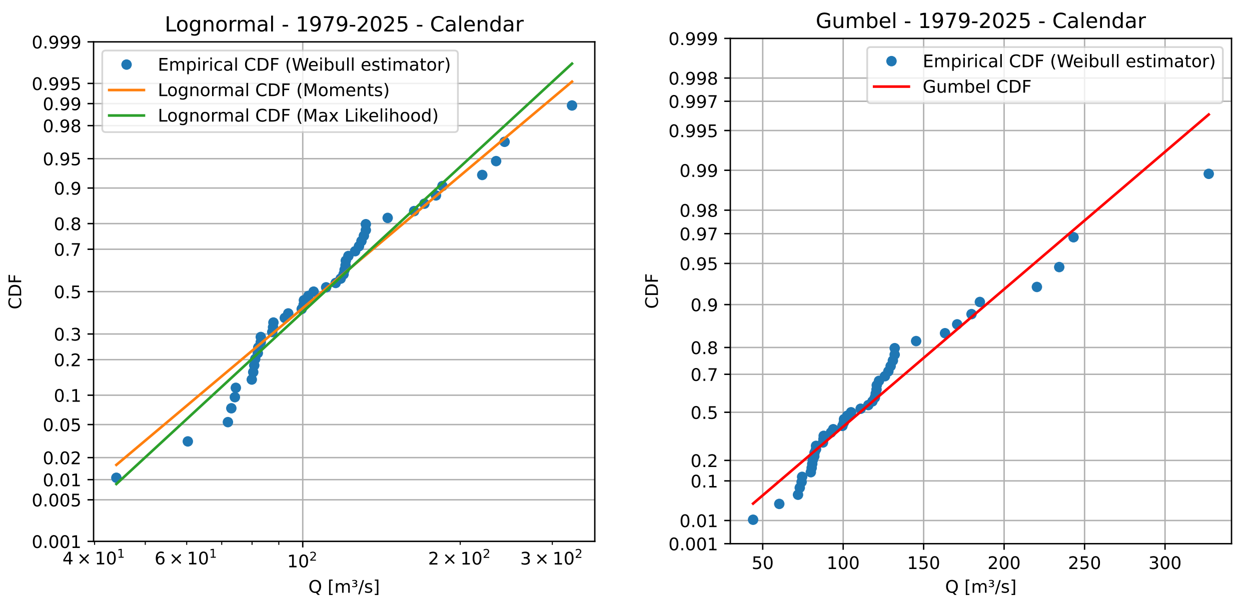


Figure 4.1: Comparison between observed and theoretical distributions (Lognormal, Moments & Maximum Likelihood methods, Gumbel).

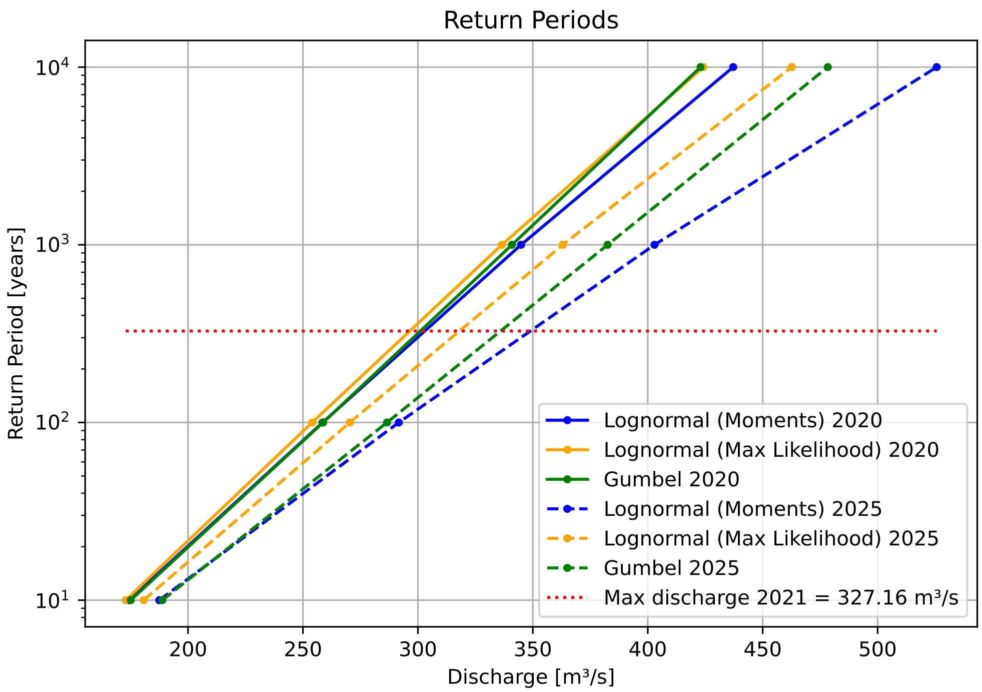


Figure 4.2: Characteristic discharges for return periods between 10 and 10.000 years.