LGCIV 2055 Analysis and mitigation of floods – Assignment 1 - Statistical methods

Statistical analysis of discharges on the Ourthe River

You are asked to perform a statistical analysis of the annual maximum discharges on the Ourthe River (Belgium) based on data from the Hydrology Office of the Walloon Region (data available at: https://hydrometrie.wallonie.be/home/observations/debit.html?mode=table). Consider the data for the Hotton station from 1 January 1979 to 16 September 2025. A file containing the average daily flows is available on Moodle.

1. Analysis over the years 1979-2020

You are first asked to determine the extreme flows for several return periods (see Table 2 as an example) using the data from 1 January 1979 to 31 December 2020. The distributions to use are those seen during the classes:

- the two lognormal distributions (moments + maximum likelihood);
- the Gumbel distribution.

The statistical analysis will include:

- the graphical illustration of the three distributions (linear relationship between the horizontal and vertical axes) together with the observed data points estimated using the Weibull estimator (see Table 1);
- a test of the validity of the proposed distributions using Kolmogorov-Smirnov and Chi-squared (4 and 5 classes) tests;
- a summary table of the parameters, extreme flows and hypothesis tests as illustrated in Table 2.

X_i	$F_X(X_i)$	$(i-0.5)/n$	i/n	$F_X(X)- i/n$	X_i	$F_X(X_i)$	$(i-0.5)/n$	i/n	$F_X(X)- i/n$
50.2	0.000	0.017	0.035	– 0.035	57.5	0.455	0.535	0.552	– 0.097
50.2	0.000	0.052	0.069	– 0.069	58.5	0.498	0.569	0.586	– 0.088
50.3	0.008	0.086	0.103	– 0.095	59.9	0.553	0.603	0.621	– 0.068
51.8	0.124	0.121	0.138	– 0.014	60.8	0.585	0.638	0.655	– 0.070
51.9	0.132	0.155	0.172	– 0.041	61.6	0.612	0.672	0.690	– 0.078
51.9	0.132	0.190	0.207	– 0.075	61.8	0.618	0.707	0.724	– 0.106
52.1	0.146	0.224	0.241	– 0.095	61.9	0.621	0.741	0.759	– 0.137
52.2	0.153	0.259	0.276	– 0.123	69.4	0.797	0.776	0.793	0.004
52.4	0.167	0.293	0.310	– 0.143	69.9	0.805	0.810	0.828	– 0.022
52.7	0.187	0.328	0.345	– 0.157	72.7	0.846	0.845	0.862	– 0.017
53.7	0.252	0.362	0.379	– 0.127	72.8	0.847	0.879	0.897	– 0.050
54.5	0.300	0.397	0.414	– 0.114	74.9	0.871	0.914	0.931	– 0.060
55.4	0.351	0.431	0.449	– 0.098	87.3	0.954	0.948	0.966	– 0.012
55.9	0.377	0.466	0.483	– 0.106	145.0	1.000	0.983	1.000	– 0.000
55.9	0.377	0.500	0.517	– 0.140					

Table 1 – Example of the analysis of a cumulative distribution function
(Daily rains in Taza – Exponential distribution).

	LN moments	LN maxi likelihood	Gumbel moments
Parameters	$m_{\ln Y} = 6.535$ $\sigma_{\ln Y} = 0.288$	$m_{\ln Y} = 6.531$ $\sigma_{\ln Y} = 0.313$	$u = 623.13$ $\alpha = 0.00608$
Extreme discharges			
$T_r = 10$ years	996 m³/s	1025 m³/s	993 m³/s
$T_r = 100$ years	1346 m³/s	1421 m³/s	1380 m³/s
$T_r = 1000$ years	1677 m³/s	1805 m³/s	1760 m³/s
$T_r = 10000$ years	2009 m³/s	2197 m³/s	2138 m³/s
K-S test	$\lvert -0.098 \rvert < 0.214$	$\lvert +0.083 \rvert < 0.214$	$\lvert +0.087 \rvert < 0.214$
Chi squared			
4 classes	0.75 < 3.84	0.75 < 3.84	1.75 < 3.84
5 classes	0.50 < 5.99	2.06 < 5.99	1.44 < 5.99
6 classes	3.62 < 7.81	3.62 < 7.81	2.87 < 7.81

Table 2 – Example of results summary: parameters, extreme discharges and hypothesis tests.

For this analysis, you are asked to consider the calendar years (complete or incomplete) as well as the hydrological years, and to comment on the results.

2. Analysis over the years 1979-2025

Considering the catastrophic floods that occurred in the Ourthe basin during the summer 2021, you are asked to evaluate numerically and comment the influence of these floods on the characteristic discharges (Q_{10}, Q_{100}, etc.) and on the return period for the maximum flood observed in 2021 (Table 3). Don't forget to clearly specify this maximum flood discharge in your report. Consider only the calendar years for that analysis.

		Q_{10} (m³/s)	Q_{100} (m³/s)	Q_{1000} (m³/s)	Q_{10000} (m³/s)	*Maximum discharge in 2021* T_r (years)
Hotton 1979 - 2020	LN mom LN ml Gumbel					
Hotton 1979 - 2025	LN mom LN ml Gumbel					

Table 3 – Comparison table.

All the results should be presented and discussed in a report (5 pages) to be delivered on Moodle. Your calculations can be made with a spreadsheet or Python using data-driven packages as pandas, numpy and scipy (preferred option). Be accurate and make relevant conclusions when appropriate. Think about the usefulness of such an analysis and give an interpretation of your results consequently. Your report should be comprehensive and include the key results in graphical form. Your spreadsheet or Python code should be delivered with the report and must be clear enough to understand where your mistakes come from (if any).