Assignment 1 – Detailed instructions

Hydrology for engineer 2021-2022

Items marked **in bold** must be included in the report.

1 Process rainfall data from MeteoSwiss

Specific goal: import the data from file `data.txt` and compute the annual maxima for durations D of: 1, 3, 6, 12, 24, 48 hours.

1. Data inspection: open the raw file and metadata to understand how the file is organized. For example, the code `rre150h0` stands for "mm of water cumulated over one hour". How many headerlines does the data file have? How is missing data reported by MeteoSwiss? This is important information that will be used to import the data.

2. Data import: import the data into a matlab table using the command `readtable`. A working example is already provided in the template. Check the options for the headerlines and for the empty values.

3. Data cleaning: extract the time and rainfall depth into vectors t and h. Convert the dates into matlab timestamps using `datetime` (this will take some good 20 seconds). Generate two vectors `m` and `y` with the month number and year of each date (see functions `month` and `year` applied to a date). For the sake of simplicity, replace the missing values (that appear in Matlab as `NaN` – not-a-number) with zero. To do so, you can use the function `isnan`, which returns a `true` (or 1) for every `NaN` and a `false` (or 0) otherwise.

4. **Plot with yearly rainfall**: Compute the total annual precipitation over each year. Then, show it in a plot: on the x-axis there should be the years and on the y-axis the total annual precipitation (in mm).

5. Compute rainfall maxima of a certain duration: if you have a series of hourly rainfall as

$$h\,[mm] = \{0, 2.3, 4.0, 0, 5.6, 2.8, 6.9, 0, 3.1, 0\}$$

   how could you compute the maximum precipitation over 3 hours? The maximum precipitation of 3 hours is the maximum value over 3 consecutive hours. There are multiple ways to implement this. One way is to create a moving window of 3 elements. 'Moving' windows means that you take the elements `1:1+2`, then `2:2+2`, `3:3+2` etc until you reach the length of the record (which is 10) `10-2:10-2+2`. Then, you can compute the sum of the precipitation values over each moving window and retain the maximum. *[You don't need to include this in the report].*

6. Compute rainfall maxima of a certain duration over a year: Let's extend to a yearly duration. You first need to select data belonging to one year using again the vector `y` (created using the function `year`). For example, `y == 2000` will return a logical vector with `true` (or 1) on all timestamps belonging to year 2000 and `false` (or 0) otherwise. The 8784 rainfall datapoints belonging to year 2000 (which is a leap year, otherwise they would be 8760) are quickly obtained through logical indexing `h(y == 2000)`. *[You don't need to include this in the report].*

7. Extend to multiple years and multiple durations: if you managed to do the previous steps, you can quickly add external `for` loops to compute and store the rainfall maximum over each year of data and for any duration you need (as listed on top, here you want durations D of: 1, 3, 6, 12, 24, 48 hours). It is good to preallocate a matrix to store the maxima, with one row per year and one column per duration. Call this matrix 'AnnualMax'.

8. Save your results for the following parts of the assignment: the matrix you produced will be used in the next parts of the assignment. You can save variables to a matlab file through the command `save`. For example, `save assignment1_output_part1 var1 var2` (and you can add more) saves the variables `var1` and `var2` to the file 'assignment1_output_part1.mat'. You should save the matrix `AnnualMax` and the durations `D`.

2 Fit a Gumbel distribution and calculate critical rainfall depths

Specific goal: fit a Gumbel distribution to the precipitation maxima and compute the critical rainfall corresponding to return periods $T = 10, 40, 100$ years.

1. Rainfall empirical frequency: Load the rainfall maxima computed in Part 1 and compute the empirical frequency F_h of the rainfall maxima using the Weibull plotting position formula. Also, compute the reduced variable Y_F. If you want, you can already plot your empirical frequencies F_h against precipitation depths h.

2. Fit the Gumbel distribution: now for each rainfall duration you need to fit the Gumbel distribution to the empirical frequency. Use the two methods seen in class (method of moments and Gumbel method) to determine the parameters α and u. Remember the formulas depends on the mean and the standard deviation of h and Y_F).

3. Compute analytical Gumbel distributions: use the parameters α and u to implement the Gumbel distribution for each rainfall duration.

4. **Plot**: plot the fitted Gumbel distribution against rainfall depth for each duration (use continuous smooth lines). To obtain a smooth curve you should plot it for a large number of h values, not just the few h that were measured. On the same plot, add the empirical frequency of the yearly maxima computed through the Weibull plotting position (use points).

5. T_h and h: Use the Weibull plotting position to compute the return periods T_h associated with the measured events h. Also, invert the Gumbel distribution and compute the return period as a function of rainfall depth.

6. **Plot**: Plot the estimated rainfall depth obtained through the Gumbel distribution against the return period, for each rainfall duration. Then, add the measured rainfall depths against the empirical return periods. Again, use points for the empirical frequency and smooth lines for the Gumbel distribution (to obtain a smooth curve you should plot it for many T values, not just the few T that were measured).

7. **Include a table**: Create a matrix `H_Gum` with the estimated rainfall depth for return periods of: 10, 40 and 100 year (rows) and for each duration (columns). Include this matrix as a 3x6 table in your report (no need to do the table in Matlab).

8. Save your results for the following parts of the assignment: Save the matrix `H_Gum`, the return periods `T` and the durations `D` to a matlab file named 'assignment1_output_part2.mat'. This file needs to be loaded at the beginning of the Assignment part 3.

3 Construction of DDF curves

Specific goal: build DDF curves of the type:

$$h = \frac{c\,D}{D^e + f} \tag{1}$$

where h is the estimated rainfall depth, D is the event duration and c, e, and f are parameters that need to be estimated for each return period T.

1. Parameter calibration: Implement a "brute force" algorithm to calibrate the parameters c, e and f of the DDF curve for each return period. Look for the best fit in these parameters ranges:
   - $c = [0, 100]$
   - $f = [-1, 1]$
   - $e = [0, 1]$

   Hint: implement nested loops in order to browse the entire domain of the parameters `c`, `e` and `f`. At each loop, compute the value of the rainfall depth estimated by the current parameter combination. Then, compare it to the estimated rainfall depth from the Gumbel distributions (as obtained from Assignment 1 Part 2). Use the sum of the squared differences between the computed value and the Gumbel estimate as a measure of the error. You have to browse the entire domain and retain only the parameter set that provides the smallest error. Start by exploring just few parameter values (e.g. `c=(0, 50, 100)` until you are sure that your code is right. Then, explore a larger number of parameter values until you obtain a sum of squared errors smaller than:
   - 25 for $T = 10\ yr$
   - 60 for $T = 50\ yr$
   - 120 for $T = 200\ yr$

2. **Include a Table**: present in a table the calibrated parameter values and the error obtained for each return period.

3. **Plot**: show the DDF curves for each return period in a single figure. Also include the Gumbel estimates for each duration. As usual, to obtain smooth DDF curves you should compute them on a large number of rainfall durations D (much larger than the 6 measured durations).

4 General questions

1. How could you compute, for every year, the number of days without any rainfall? *only reply in words*

2. In this project you used the Gumbel distribution to approximate the empirical data. How can you evaluate the goodness of this approximation? What can you do in case the approximation is unsatisfactory?

3. You were asked to provide an estimate of the maximum rainfall depth with a return period of 100 years, even if you only had 39 years of data (extrapolation). How does your estimate of this 100-years event change if the α parameter is affected by a $\pm 10\%$ uncertainty?

4. You want to design a structure at the site where you computed the DDF curves. The structure needs to resist to critical rainfall events up to 80 mm in 16 hours. Every how many years, on average, will the structure fail to contain rainfall events of 16 hours? *there are multiple ways to answer this question*