Assignment 1 final report
Hydraulic Loading
CIEM3220 Module B2-2 - River Functions and Managements

Group 5
Lucas Terlinden-Ruhl 5863937
Antonio Magherini 5838215

Faculty of Civil Engineering and Geosciences
University of Technology Delft
The Netherlands
May 24, 2023

1 Introduction

The site is in the Colorado River and is located in Grand County near Cisco and downstream the Grand Junction, in the Utah State. The gauge collects daily data of flow discharge Q and daily suspended sediment discharge q_{ss} since 01/05/1949 (`https://waterdata.usgs.gov/nwis/inventory?agency_code=USGS&site_no=09180500`). For this assignment, mean daily flow discharge data from 1949 to 2022 have been used, while for the suspended sediment discharge data records stop at 1984.
The site is located in the so-called Upper Basin of the River, which is mainly fed by snow and rain precipitations during winter (`https://en.wikipedia.org/wiki/Colorado_River`). Therefore, we expect that extreme events are linked to the amount of precipitation that falls in the basin of the Colorado River during autumn and winter season and to snow and ice melt in late spring and early summer season.

2 Methods

To begin with, data units have been transformed from US to SI units.

Both flow and suspended sediment discharge data have been plotted as time series and then as a bivariate scatter plot. Consequently, an histogram of both parameters has been plotted to show the density of relative occurrence depending on the magnitude. Annual maxima of flow discharge have been sampled using the block maxima method. With this outcome, a GEV and Gumbel distributions have been fitted, using the **scipy** library, and also compared to the empirical CDF.
Confidence intervals have been displayed assuming an accuracy of 95% (see more in Section 4).

QQ-plots have been included as well to assess the goodness of the fit.

A comparison between the GEV, the Gumbel and the empirical CDF have been made in terms of Return Period. The x-axis has been extended in order to extrapolate data for future predictions for both GEV and Gumbel distributions (see figure 3).

To check the correlation between the two variables, the discharge dataset has been reduced to match the same time window as the suspended sediment discharge dataset (1949-1984). In this way the covariance Cov and the Pearson correlation coefficient ρ of daily mean data have been computed.

In order to further investigate the correlation between the two variables, weekly mean, week and yearly maxima data have been sampled as well and ρ has been computed again.

3 Results

Flow discharge results in almost one peak per year as shown in Figure 1. Suspended sediment discharge seems to have a similar evolution in time, although the peaks don't occur at the same moment in time.

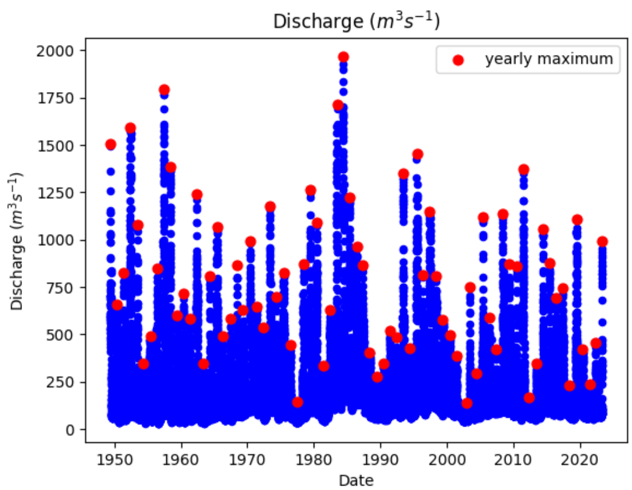

Figure 1: Time series of flow discharge annual maxima (red dots).

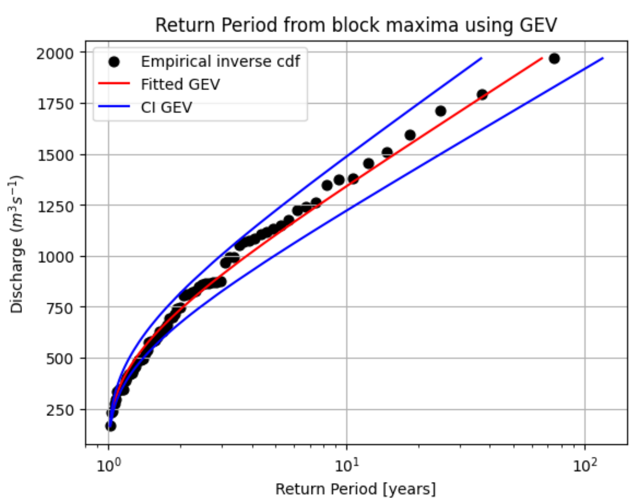

Figure 2: 95% Confidence intervals for GEV fit.

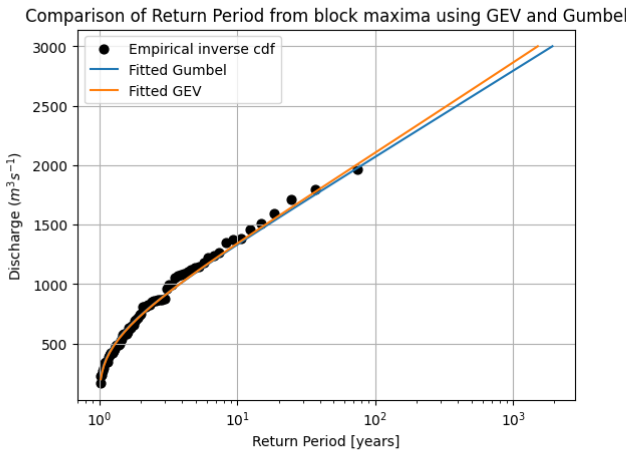

Figure 3: Return period comparison using data extrapolation with both GEV and Gumbel distributions.

Figure 2 shows the large uncertainty associated with fitting an extreme value curve to a data set that contains a small record. Figure 3 shows that both extreme value curves are quite similar.

Numerous maxima of suspended sediment discharge seems to coincide with relatively small discharges. Moreover, these maxima don't seem to have a temporal correlation (i.e., sudden increase).

The small correlation between discharge suspended sediment discharge can be seen when considering the Pearson correlation coefficient of the bivariate dataset $\rho = 0.639$. A stronger correlation can be achieved by looking at the weekly means as seen in Figure 4, which also demonstrates a temporal relationship. From a physical point of view, this relationship can be described with a hysteresis behavior, for which the largest sediment discharge does not occur together with the largest value of flow discharge but rather with a slightly smaller one. Again, when looking at Figure 4, for the sediment discharge an increasing trend over the weeks can be identified and the peak is reached around week 25 (end of May), a few weeks before the flow discharge peak.

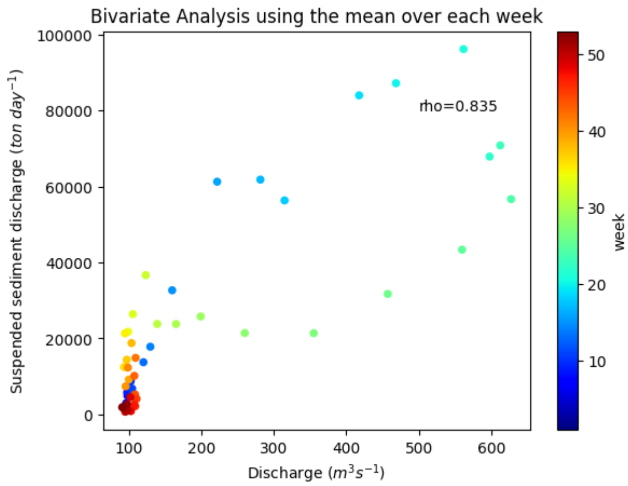

Figure 4: Bivariate Analysis using the mean values of each week of the year.

4 Discussion

The first consideration that can be done regarding flow discharge peaks is the occurrence during the year. In fact, as it can be retrieved when looking at the **yearly_maxima_discharge** dataframe (see attached Jupyter Notebook file), the large majority of the peaks occur in late spring/early summer (specifically during the first half of June). These high discharges are related to snow and ice melt contribution.

For the GEV fit, data had to be initially sorted by increasing values and for it to be effectively fitting the data the first two data points had to be deleted, as the outcome was nonsensical. The same procedure has been applied for the Gumbel distribution, even though in this case getting rid of the first two data points was not necessary for having a good fit, but was performed for consistency.

As it is possible to see from figure 3, GEV and Gumbel fit are pretty close. This is because the GEV is of type 1 and has a ξ value that is close to zero (which means it is part of the Gumbel family)

From a physical point of view, the suspended sediment discharge peaks should occur together with the flow discharge peaks. However, this is not completely true as we can see from the bivariate analysis of daily values and corresponding ρ value. In this case, further analysis may be helpful to understand what caused such huge sediment discharge peaks (mining activities, river interventions, landslides that brought sediments into the river).

The confidence intervals have been computed using a value of confidence of 95%. Moreover, the precision of the results was chosen to be equal to 5%. The reason behind the choice of this value is related to the type of data we have used. Without further specific information, we looked at the data file we had downloaded. As it can be seen, almost all flow and suspended sediment discharge data are labeled as "A (Approved for publication – Processing and review completed.)" (see `https://help.waterdata.usgs.gov/codes-and-parameters/instantaneous-value-qualification-code-uv_rmk_cd`). Therefore, from the "Discharge Measurement Quality Code" webpage (see `https://help.waterdata.usgs.gov/codes-and-parameters/discharge-measurement-quality-code`), we assumed that our data were of type "G (Good).

Finally, figure 4. The hysteresis behavior causes the sediment discharge to reach its peak (end of May) a couple of weeks before the peak of the flow discharge. During winter, ice and snow cover limits the erosive process of the basin hill slopes. Large temperatures in spring cause the ice and snow cover to melt, causing a large seasonal change. Only one-quarter of the weeks have a flow discharge that is larger than 100 $m^3\ s^{-1}$ and they all occur in late spring, early summer

5 Conclusion

The extreme values were sampled using block maxima. As was seen in figure 1, this is not a limitation as the river clearly exhibits a yearly seasonality with regular maxima each year in the spring. The assumption linked with this approach is that all samples are i.i.d (independent and identically distributed).
In general, the sources of uncertainty can be various. The main ones are the accuracy of the records and the chosen statistical distribution.
Although both curves have almost the same tail, it would be recommended to use the GEV. This is because for large return periods, it is predicting larger values of discharge, than the Gumbel one (more conservative) (Figure 3).
For $RT < 20$ years, there is a large confidence in the fit, and the value of the GEV can be used. For larger RT, ($20 < RT < 70$), uncertainties increase, which means it can be up to interpretation which value (within confidence interval) should be chosen for risk assessment analysis. Finally, for events with large RT (not recorded), extrapolation brings about large amounts of uncertainty. It would be recommended to use a process like GRADE to simulate meteorological conditions that could lead to high discharges events to increase the duration of the record.
Finally, an argument could be done relating to the accuracy of the suspended sediment discharge measurement, specifically the extreme values. A very large difference between the maxima and the rest of the values is observed. This fact made us believe that outliers could be present in the data.
Assuming the device was precise, it would be recommended to find the source of the sediment as both variables seem to show a hysteresis behavior.