Flood frequency analysis of Casimcea river

C Cerneagă¹ and C Maftei²

¹ Ovidius University of Constanta, Doctoral School of Applied Sciences, Constanta, Romania
² Transilvania University of Brasov, Faculty of Civil Engineering, Brasov, Romania

* constantin.cerneaga@gmail.com

Abstract. Flood frequency analysis is the most important statistical technique in understanding the nature and magnitude of high discharge in a river. This technique is used for establishing a relationship between flood magnitude and frequency of occurrence (return period). This study performs the flood frequency analysis of Casimcea river. The Casimcea catchment is located in the Dobrogea-Littoral basin of south-eastern Romania, and covers parts of two counties, Tulcea and Constanta. The main river of Casimcea catchment flows into the Tasaul lake and has two main tributaries (Ramnic and Cartal). In this study, Log-Normal, Log-Pearson Type III and Gumbel probability distributions have been applied on maximum discharge time series data of four hydrological stations, Casimcea and Cheia on Casimcea River and Pantelimonu on its main tributaries Ramnic and Cartal. For this study have been used the annual maximum series data spreading from 29 to 52 years. The frequency analysis will be performed with Hydrognomon, open-source software for the analysis of hydrological data. Hydrognomon is free software licensed under the General Public License (GPLv3).

1. Introduction
Floods occur due to various natural and anthropogenic factors and are most common environmental hazards, affecting people globally. Flooding has been a larger concern today due to rapid development in river catchment area, which increases river runoff and decreases river capacity.

A return period is known as a recurrence interval, which is an estimate of the interval of time between events like an earthquake, flood or river discharge flow of a certain intensity or size [1]. The terms "10-year", "50-year", "100-year" and "500-year" flood are used to describe the estimated probability of a flood event happening in any given years [2].

Flood frequency analysis is the procedure for estimating the frequency of occurrence (return period) of a hydrological event such as flood. The technique involves using observed annual peak flow discharge data to calculate statistical information such as mean values, standard deviations, skewness and recurrence intervals [3].

Various probability distributions are currently used to predict expected rainfall in different return periods, as rainfall varies with time and location [4]. Frequency analyses of rainfall data have been performed for different return periods [5].

The Kolmogorov-Smirnov test and Normal, Log-Normal, Log-Pearson Type III and Gumbel distributions are used to fit the annual extremes rainfall data with 10, 25, 50, 100 and 200 years return periods.
2. Material and methods

2.1. Study area

The Casimcea catchment is located in the Dobrogea-Littoral basin of south-eastern Romania (figure 1), and covers parts of two counties, Tulcea and Constanta.

In the present study 29 to 52 years of annual maximum peak discharge data was used for analysis (figure 3). The size of the sample data can be considered as reliable for projecting discharge scenarios. „Romanian Waters” National Administration has collected the data from three gauging stations, Casimcea (upstream), Pantelimonu (tributaries Cartal and Ramnic), and Cheia (downstream).

Figure 1. Study area.

Figure 2. Hydrological stations.
2.2. Methods

To predict extreme hydrological discharge using frequency analysis, Hydrognomon software [6] and EasyFit software [7] are used. The most common probability distribution functions (PDF) used for extreme discharge data are Log-Normal, Pearson Type III, Log-Pearson Type III and Gumbel distributions [8]. U.S. Water Resources Council [9] recommends Log-Pearson Type III PDF method found to be yielding good results in many applications, particularly for flood peak data. Hosking and Wallis [10] consider that Gumbel PDF procedure is a rational approach to predict the flood recurrence interval.

Kolmogorov-Smirnov (K-S) test from Hydrognomon and Anderson Darling (AD) from EasyFit are used to test if the discharge sample comes from a specific PDF. If the K-S test statistic \( D \) is greater than the critical value, the null hypothesis (\( H_0: \) the data follow a specified distribution) is rejected. Critical values of \( D \) statistic for different significance level, \( \alpha \), depend on number of observations (\( n \)) [11]. For example, at \( \alpha \) of 5%, and 50 observations, the critical value is 0.190.

The Anderson-Darling (AD) test is similar to K-S but is more powerful especially when all the data values are considered, not just the one that produce the maximum difference. The statistic of AD test, \( A^2 \), depends on number of observations (\( n \)) and the distribution function (\( F \)). If \( A^2 \) at a certain significance level, \( \alpha \), is greater than the critical values of the selected distribution function, then the null hypothesis is rejected.

3. Results and discussion

According with Kolmogorov-Smirnov (table 1), the Log-Normal PDF is not accepted for all stations investigated. Pearson Type III is rejected for Pantelimonu station located on Cartal tributary. Log-Pearson Type III is not accepted for Pantelimonu station located on Ramnic tributary. We observe that the Pearson Type III distribution for Casimcea station and Log-Pearson Type III distribution for Cheia station have the best fit with “attained \( \alpha \)” of 51.5%, and 66.6% respectively. For Pantelimonu station on Ramnic tributary the better distribution is Gumbel with “attained \( \alpha \)” of 15.74%.
Table 1. The results of the Kolmogorov-Smirnov for all stations investigated.

| Kolmogorov-Smirnov test | α = 1, 5 and 10% | Attained α (attained threshold) |
|-------------------------|------------------|---------------------------------|
| **Casimcea st. on Casimcea** | | |
| LogNormal | REJECTED | 0.23% |
| Pearson III | ACCEPT | 51.50% |
| Log Pearson III | ACCEPT | 23.28% |
| EV1-Max (Gumbel) | ACCEPT | 0.05% |
| **Cheia st. on Casimcea** | | |
| LogNormal | ACCEPT | 22.28% |
| Pearson III | ACCEPT | 14.34% |
| Log Pearson III | ACCEPT | 66.61% |
| EV1-Max (Gumbel) | ACCEPT | 13.79% |
| **Pantelimonu st. on Cartali tributary** | | |
| LogNormal | REJECTED for 1% and ACCEPT for 5 and 10% | 1.37% |
| Pearson III | REJECTED | 0.00% |
| Log Pearson III | ACCEPT | 42.53% |
| EV1-Max (Gumbel) | REJECTED | 0.01% |
| **Pantelimonu st. on Ramnic tributary** | | |
| LogNormal | REJECTED for 1% and ACCEPT for 5 and 10% | 1.44% |
| Pearson III | ACCEPT | 15.74% |
| Log Pearson III | REJECTED | 0.55% |
| EV1-Max (Gumbel) | ACCEPT | 5.11% |

Face to the situation presented below, has been performed the Anderson-Darling test under EasyFit in order to choose the best PDF. The $A^2$ statistic results (table 2) show that the better distribution function for all stations is Log-Pearson Type III because the $A^2$ value is less than Log-Normal distribution. The Gumbel distribution was rejected for all stations and for all α level of significance.

Table 2. The results perform with EasyFit.

| AD test | α = 1, 5 and 10 % | $A^2$ statistic |
|---------|------------------|----------------|
| **Casimcea on Casimcea** | | |
| LogNormal | ACCEPT | 1.047 |
| Log Pearson III | ACCEPT | 0.61 |
| **Cheia on Casimcea** | | |
| LogNormal | ACCEPT | 0.64 |
| Log Pearson III | ACCEPT | 0.3 |
| **Pantelimonu st. on Cartal** | | |
| LogNormal | ACCEPT | 1.047 |
| Log Pearson III | ACCEPT | 0.89 |
| **Pantelimonu st. on Ramnic** | | |
| LogNormal | ACCEPT | 0.44 |
| Log Pearson III | ACCEPT | 0.36 |
As we can see Log-Pearson Type III was accepted but the values for different return period presented in table 3 are very high compared to those obtained using Hazen empirical formula [12] and with the annual maximum peak discharge data that we use. Hazen formula is calculated with the following equation:

\[ P = \frac{2m-1}{2n} \]  

(1)

where \( m \) is rank assigned to the data after arranging them in descending order of magnitude and \( n \) represent number of records. Thus, the maximum value is assigned \( m = 1 \), the second largest value \( m = 2 \), and the lowest value \( m = n \).

Table 3. The maximum discharge for different return period.

| T(Max) = | 200 y | 100 y | 50 y | 25 y | 10 y |
|---------|-------|-------|------|------|------|
| Casimcea st - Casimcea river |       |       |      |      |      |
| Log Pearson Type III | 6624.2 | 2964.4 | 1269.6 | 513.3 | 136.3 |
| Hazen | - | 395.3 | 324.6 | 167.6 | 75.6 |
| Cheia st - Casimcea river |       |       |      |      |      |
| Log Pearson Type III | 4975.2 | 2635.8 | 1359.4 | 676.8 | 248.8 |
| Hazen | - | - | 379.9 | 350.3 | 270.6 |
| Pantelimonu st - Cartal tributary |       |       |      |      |      |
| Log Pearson Type III | 7432.8 | 3448.98 | 1540.67 | 654.671 | 188.877 |
| Hazen | - | 484.6 | 395.1 | 312.5 | 156.8 |
| Pantelimonu st - Ramnic tributary |       |       |      |      |      |
| Log Pearson Type III | 10132.5 | 3913.11 | 1464.76 | 526.211 | 124.026 |
| Hazen | - | 130.6 | 121.8 | 111.7 | 60.7 |

4. Conclusions
In this paper, the main purpose is to understand the frequency analysis of annual maximum discharge on Casimcea River. After frequency analysis, Log-Normal and Log-Pearson Type III are the best distribution functions which are provided by Anderson-Darling test and Log-Pearson Type III by Kolmogorov-Smirnov test. The discharge values obtained for different return period based on those theoretical distribution functions are different from the real ones.

Based on this study, any distribution function should not be considered suitable for analysis of maximum annual data of Casimcea River.

An upcoming paper will assess other distribution functions to fit the best function. The study results lead to the preparation of a flood hazard, which are essential for flood management plan.

References
[1] Rakhecha P R and Singh V P 2009 *Applied Hydrometeorology*. Heidelberg: Springer
[2] Watson D and Adams M 2010 *Design for Flooding*. Canada: John Wiley and Sons
[3] Derman W H A W, Mutalib N J A and Mukhatar N Z 2017 *Determination of return period for flood frequency analysis using normal and related distributions*. IOP Conf.: Journal of Physics
[4] Upadhaya A and Singh S R 1998 *Estimation of consecutive day’s maximum rainfall by various methods and their comparison*. Indian J. Soil Conserv. 26, 193-201
[5] Vivekanandan N 2012 *Intercomparison of extreme value distributions for estimation of ADMR*. Int. J. Appl. Engg. Technol., 2 (1), 30-37
[6] http://www.hydrognomon.org/, last accessed 2020/11/02
[7] https://easyfit.soft32.com/, last accessed 2020/11/02
[8] Amin M T, Rizwan M and Alazba A A 2016. *A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan*. Open Life Sciences, 11:432-440
[9] Hammett K M and DelCharo M J 2001 *Estimating the magnitude and frequency of floods for streams in West-Central Florida*. Scientific Investigations Report 2005-5080
[10] Hosking J R M and Wallis J R 1997 *Regional Frequency Analysis – An Approach Based on L-Moment*. Cambridge University Pres, Cambridge, UK
[11] http://www.itl.nist.gov/div898/handbook/, last accessed 2021/02/04
[12] Hazen A 1930 *Flood flows – A study of Frequencies and Magnitudes*. Jhon Wiley & Sons, Inc.