

Probability Theory Use in Hydrology (Sitnica River Shed)

Zekirija Idrizi¹, Isak Idrizi² and Kujtim Zena¹

1. Faculty of Civil Engineering and Architecture, Hasan Prishtina University, Prishinë 10000, Kosova;

2. Faculty of Civil Engineering and Architecture, Mother Teresa University, Shkup 1000, Macedonia

Abstract: The use of probability theory and mathematical statistics in hydrology has begun very late. Nevertheless, lately, the statistical methods in hydrology engineering and in economy aspects as well, have been found to be of vital importance, by successfully solving many problems considering the hydrological laws and the quantity evaluation of the many characteristics of different hydrological regimes. The use of the statistical methods should not be used formally, because it can lead to wrong conclusions, which would lead to bad dimensioning of the hydro-technical objects. The solving of many hydrological problems with statistical methods, is made possible with the aid of computational calculations. Considering the facts mentioned above, we have made on simple program, in excel, for achieving probabilistic results for different time frequencies, using the logarithmic functions taken by four different authors. This program has been used for analyzing the river bed and bed load transport of the Sitnica river shed, in the territory of Kosova.

Key words: Bed load transport, Sitnica river shed, hydrological analysis.

1. General Introduction

The nature of hydrological science, nevertheless, which is very advanced and highly applied in the world, encounters many obstacles. Many of the hydrological parameters are dependent on environmental conditions, as the topography of terrain, river flows, and climatic conditions as well. All of these parameters can not be evaluated using the deterministic approach. Rather that, today we apply probabilistic approach for obtaining the hydrological parameters as it is the maximum river flow (which concerns us as the ultimate parameter), and many other climatic and hydrological parameters.

The use of probability theory and mathematical statistics in hydrology began very late. Nevertheless, lately, the statistical methods in hydrology engineering and in economy aspects as well, have been found to be of vital importance, by successfully

solving many problems considering the hydrological laws and the quantity evaluation of the many characteristics of different hydrological regimes. The use of the statistical methods should not be used formally, because it can lead to wrong conclusions, which would lead to bad dimensioning of the hydro-technical objects. The solving of many hydrological problems with statistical methods, is made possible with the aid of computational calculations.

In this paper, we are analyzing the probability theory and some theoretical mathematical functions for evaluating the maximum discharge of the Sitnica river (Kosova), for different return periods of time.

2. Overview of the Sitnica River

The Sitnica water-shed (Fig. 1), mainly consists the canter part of Kosova region. Sitnica is the main branch of the Ibër river. It is distinguished not only by its watershed size (2,861 km²), but also with its mean water annual discharge (16.6 m³/s). Near the Rabovë

Corresponding author: Isak Idrizi, doctor of science, assistant professor; research fields: earthquake engineering and structural engineering. E-mail: isak.idrizi@unt.edu.mk, isak_idrizi@hotmail.com.

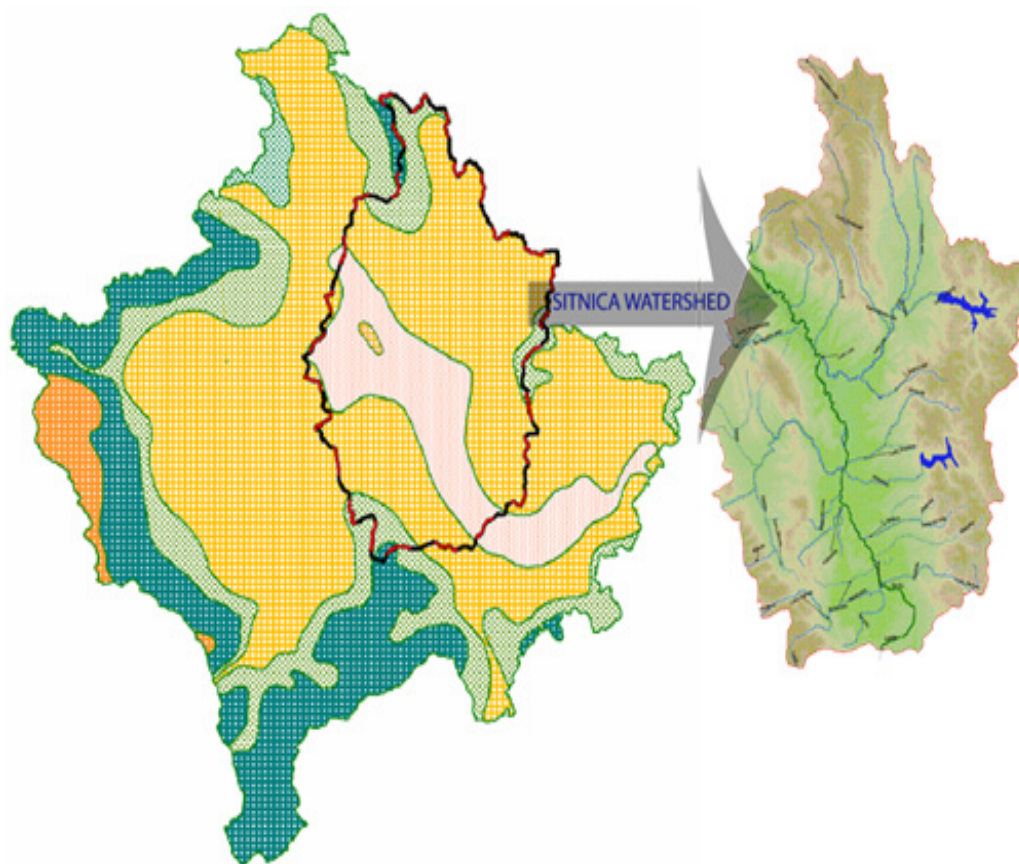


Fig. 1 Kosova region and Sitnica watershed.

village coincide the river Maticë and Shtime, thus forming the Sitnica river flow, which from Rabovcë until its pouring in the Ibër river, passes a distance of 92.3 km. With a mean longitudinal slope of 0.054% is characterized as a valley river, flowing in the alluvial valley. The highest parts of the river have the altitude of 1,750 m, while near the Ibër river, the Sitnica river has altitude of 520 m.

The south boundary of the watershed is consisted of the Sharr mountains, while the west boundary consists of the mountains of Nerodime and Drenicë. The north boundary of the watershed is consisted of the Kopauniku mountains while the east boundary is consisted of the Gollak mountains.

In order to make a more reliable forecasting of the rainfalls, mean river discharge and maximal river discharge, it is necessary to take in to account many parameters, such as the climatic characteristics, geologic and hydrologic characteristics.

In this paper, applying the probabilistic statistical calculations, we used the annual measurements of discharge of Sitnica river as minimal mean and maximal discharge. These measurements are taken from the “Nadakoc” hydrological station.

This station is near the pouring of the Sitnica river in the Ibër river, and consequently the overall discharge of the Sitnica river and its branches pass through this station point. Therefore, this station is of significant importance for the hydrometrical measurements for the overall discharge determination of the Sitnica river.

3. Hydorogical Analysis

The measurements that we have used for hydrological calculation, consist the period from the year of 1952 until 1978. These measurements are shown in Table 1.

As a first step, which aims to determine the probabilistic

Table 1 Minimum, mean and maximum river discharge taken from Nadakoc station.

No.	Hydrological year	Q_{min} (minimum discharge)	Q_{mes} (mean discharge)	Q_{max} (maximum discharge)
1	1952	0.900	9.947	34.500
2	1953	1.650	9.283	34.500
3	1954	0.370	14.073	43.000
4	1955	4.180	31.851	84.500
5	1956	0.377	20.430	56.500
6	1957	1.180	9.022	32.400
7	1958	0.169	18.600	90.200
8	1959	2.260	7.583	21.900
9	1960	0.881	14.933	54.300
10	1961	0.370	11.673	75.800
11	1962	0.608	15.446	55.300
12	1963	2.830	23.934	115.000
13	1964	3.560	8.094	13.100
14	1965	1.430	14.758	48.800
15	1966	0.950	11.297	42.800
16	1967	0.840	6.072	13.500
17	1968	0.820	4.846	15.600
18	1969	1.330	10.713	39.200
19	1970	2.020	15.353	50.300
20	1971	2.140	9.861	31.600
21	1972	1.840	8.621	25.000
22	1973	2.550	17.069	53.100
23	1974	2.260	13.927	42.400
24	1975	1.820	8.572	17.100
25	1976	4.820	21.526	63.900
26	1977	2.090	12.414	58.400
27	1978	2.120	12.417	36.200

river discharge for different return periods, is the determination of the main statistical parameters of the measurement string and the discharge distribution curve. The main statistical parameters are separated in three groups:

(1) grouping parameters:

- mean value, \bar{X} ;
- median, Me ;
- mode, Mo ;
- quintiles;

(2) spreading parameters:

- amplitude;
- dispersion (variance), $D(\sigma^2)$;
- mean quadratic deviance (σ);
- variance coefficient Cv ;

(3) shape parameters:

- asymmetric coefficient Cs ;
- excess Ek .

Considering the measurements of the Nadakoc station (maximal annual discharge), in the following, we present the main statistical parameters in Table 2 [1-3].

For graphical representation of the distribution density curve of discharge frequencies, the maximal discharges taken from Table 1 are separated in classes (intervals) as in Table 3.

Considering Table 3, in the following, we show the histogram of string frequency density as well as the reciprocal cumulative frequency.

Fig. 3 clearly shows that for higher values of discharge, the probability of exceedance of discharge is lesser.

Table 2 Statistical parameters obtained from the Nadakoc station measurement.

Statistical parameter	Formula	Obtained value considering Table 1
Mean \bar{X}	$\bar{X} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \cdot \sum_1^n x_i$	46.256 m ³ /s
Median Me	$M_e = l_e \cdot \frac{\left(\frac{1}{2} \sum f - Q\right)}{f_e} \cdot C_e$	42.8
Mode M_0	$M_o = l_0 + \frac{f_{0-1} - f_0}{f_{0-1} - 2f_0 + f_{0+1}} \cdot C$	34.5
Amplitude A	$A = X_{\max} - X_{\min}$	101.9
Mean quadratic deviance σ	$\sqrt{\frac{\sum_1^n (x_i - \bar{X})^2}{(n-1)}}$	24.535
Variance coefficient C_v	$C_v = \sigma / \bar{X}$	0.530
Asymetrical coefficient C_s	$C_s = \frac{1}{\sigma} \sum (x_i - \bar{X})^3 / n$	0.857

Note: x_i —string member;

l_e —lower boundary of the string interval consisting the median;

Q —frequency sum of the string interval which are before the string interval consisting the median;

f_e —string interval frequency consisting the median;

c_0 —string interval band consisting the median;

l_0 —lower boundary of the string interval consisting the mode;

f_{0-1} —neighbouring string interval frequency, on the left side of the string interval consisting the mode;

f_0 —string interval frequency consisting the mode;

f_{0+1} —neighbouring string interval frequency, on the right side of the string interval consisting the mode;

C —string interval band consisting the mode.

Table 3 Discharge classification in string intervals and determination of string classes frequencies.

Discharge interval (m ³ /s)	Number of occurrence in string	Frequency (Fig. 2)	Cummulative occurence	Cummulative frequency	Reciprocal cumulative frequency (Fig. 3)
13-22	5.000	18.518	5.000	18.518	100.000
22-31	1.000	3.703	6.000	22.222	81.481
31-40	6.000	22.222	12.000	44.444	77.777
40-49	4.000	14.814	16.000	59.259	55.555
49-58	5.000	18.518	21.000	77.777	40.740
58-67	2.000	7.407	23.000	85.185	22.222
67-76	1.000	3.703	24.000	88.888	14.814
76-85	1.000	3.703	25.000	92.592	11.111
85-94	1.000	3.703	26.000	96.29	7.4074
94-103	0.000	0	26.000	96.29	3.703
103-112	0.000	0	26.000	96.29	3.703
112-121	1.000	3.703	27.000	100	3.703

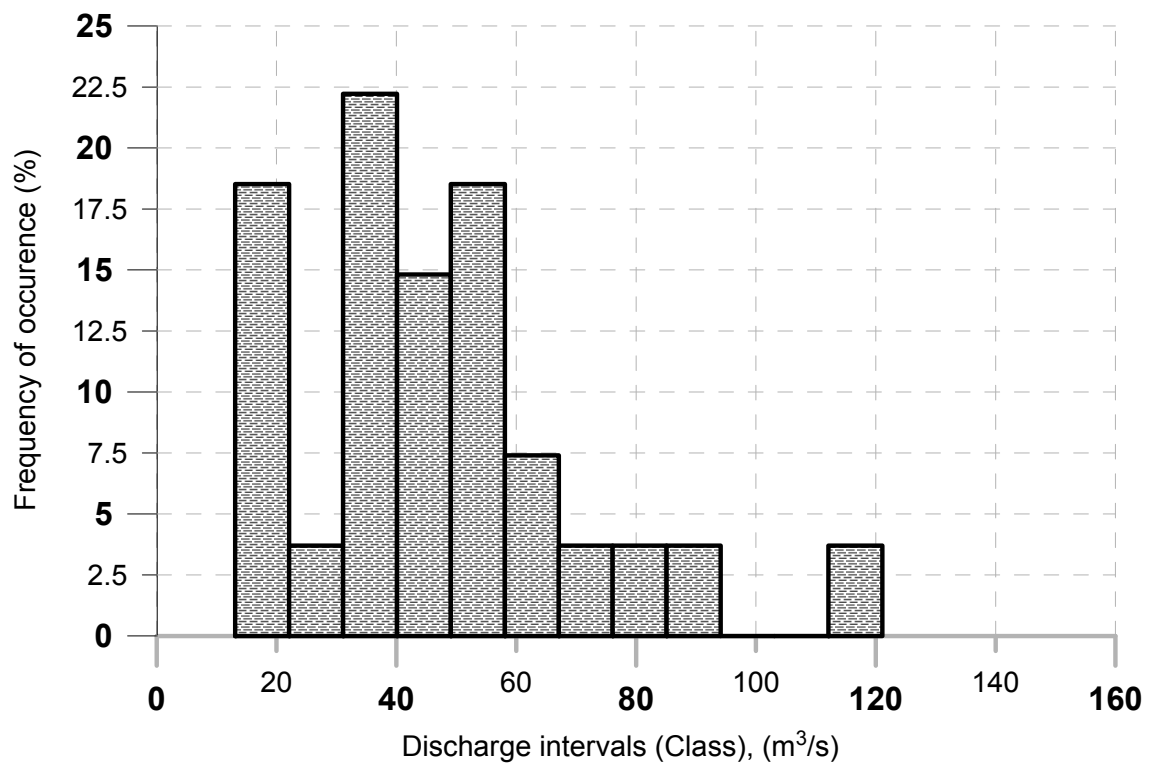


Fig. 2 Histogram of string frequency density.

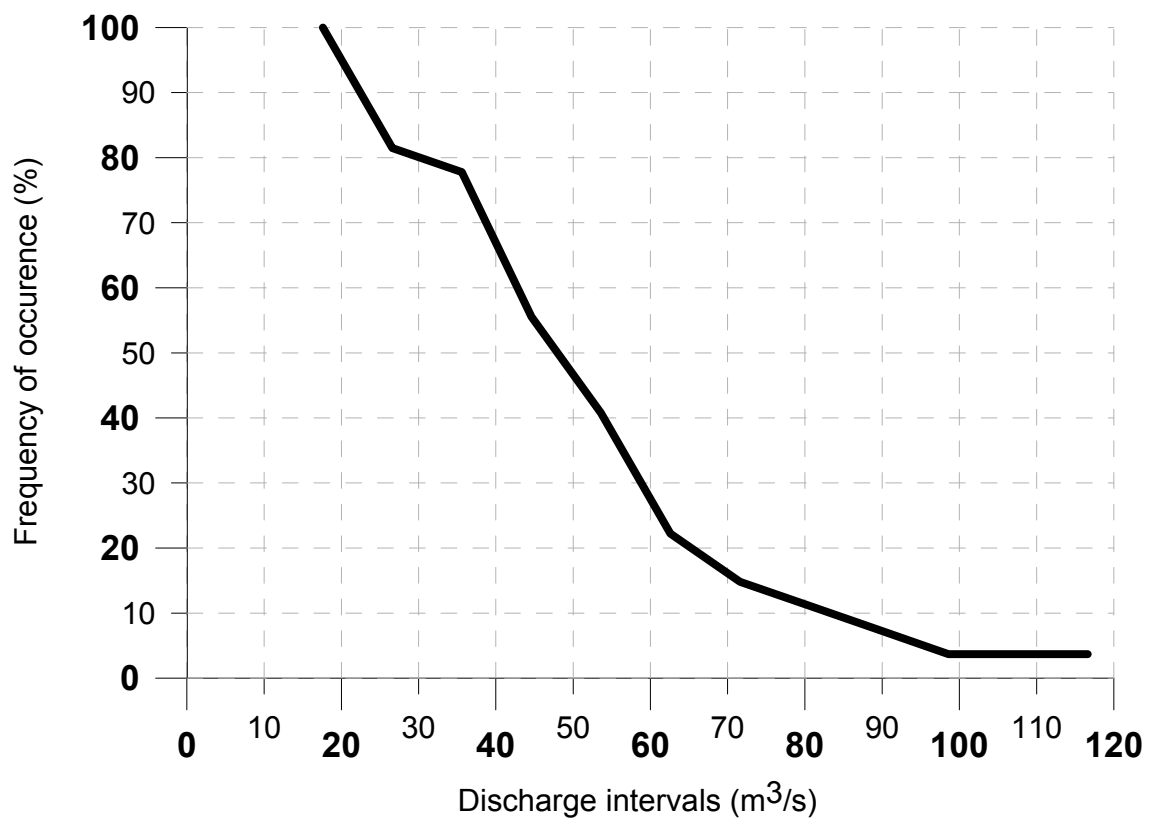


Fig. 3 Reciprocal cumulative frequency curve.

The purpose of discharge measuring and application of statistical methods for discharge measurement elaboration is to be able to forecast a more reliable estimation of the mean and maximal discharge for different return periods of time. The relation between the probability of exceedance p (%) and mean return value N is expressed by:

$$p = 1 - e^{-1/N} \text{ or}$$

$$N = -1/\ln(1 - p)$$

This correlation, for some values of probability of exceedance is shown in Table 4.

Based on Fig. 3, by transformation of the ordinate from frequency to mean return period N , is obtained Fig. 4.

From Fig. 4, it can be clearly seen that for discharge of greater value, the mean return period is greater.

In predicting and taking preventive measures against river floods, it is of great importance to know

the return period of an event (maximal discharge), usually for a return period of 100, 200 or 500 years.

Maximal discharge determination expected for long return periods of 100 or 500 years, considering only the factual measurements, would not be a rational solution and it would lead to wrong conclusions.

The solution of this problem by many authors was searched in the application of theoretical distribution function, from which would be available to further process the factual data by illuminating the more characteristic properties of the measurements string, while leaving aside some occasional deviations that occur on the distribution curve of the measurements string, Fig. 5. The application of these functions enables us a more reliable forecasting of the mean and maximal river discharge for longer return periods of time. The exactness of this solution would be directly

Table 4 Relation between the probability of exceedance p (%) and mean return value N .

p (%)	1	5	10	20	40	50	60	80	90	95
N	100	19.5	9.5	4.5	1.96	1.44	1.09	0.622	0.434	0.333

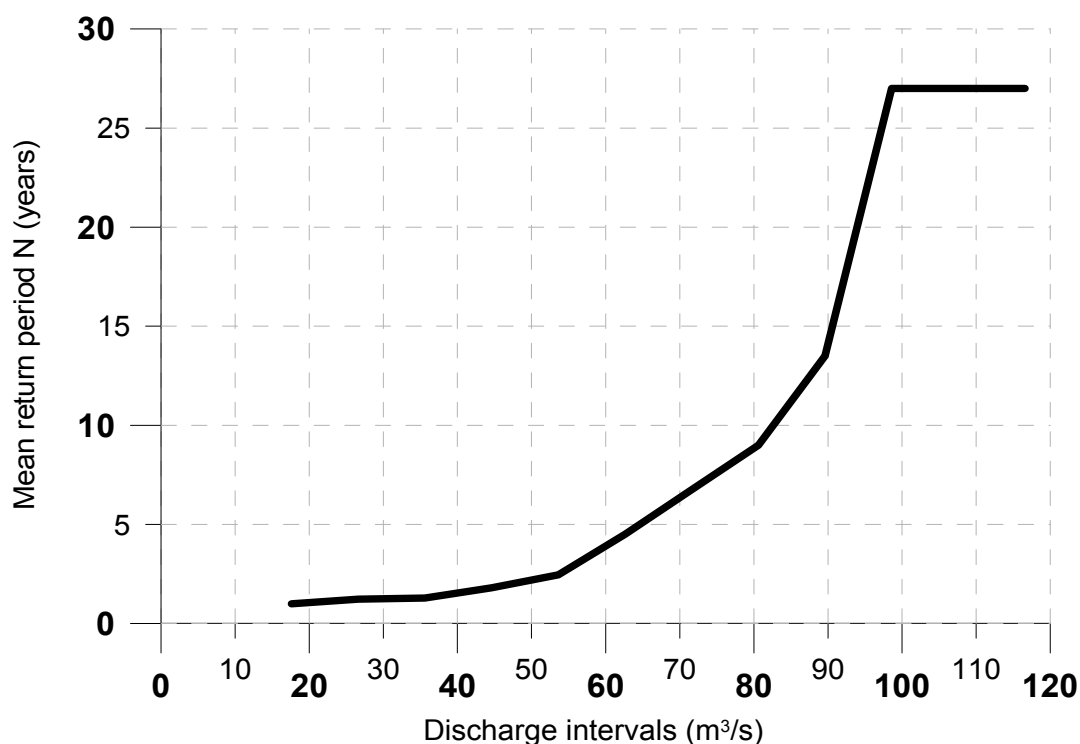


Fig. 4 Mean return value for maximal discharge (in years).

dependent on the exigency of the factual measurements as well as the length of the measurements string.

There are many theoretical distribution functions, but generally, they are classified in three major groups:

(1) Normal theoretical distribution functions, which, considering the fact that they are symmetrical, have limited use in hydrological parameter analysis. However, they can be applied for the analysis of mean annual hydrological parameters. Normal theoretical distribution function, in many literatures, is found as Gauss distribution. By modification, i.e., logarithmic transformation of hydrological parameter string, we derive more appropriate shapes of distribution curves for hydrological analysis, and their applicability can be broadened for maximal hydrological parameters as well. Such functions can be named two-parametric and three-parametric logarithmic distribution functions. Normal distribution is one of the most widespread and studied functions in the sphere of statistical mathematics. In hydrology, the application of this distribution function is executed for the fitting of the symmetrical type of factual data (experimental measurements), for obtaining of the density distribution curve while analysing the errors probability, and for other statistical results. Because of the advancements that has this function, as a result of numerous studies of the normal distribution, many researchers give priority to distribution functions derived from the normal Gaussian distribution;

(2) Asymptotic (exponential) distribution functions have found a wide applicability in elaboration of maximal hydrological parameters. Two-parametric exponential distribution Gumbel function takes part in the most applicable group of functions in hydrology. Gumbel distribution function has some advancements such as: its two-parametric distribution function, the high simplicity of determination of the function parameters, as well as the simplicity of graphical representation of this function. On the other hand, this

function has two defectives such as: the definition interval is $(-\infty, \infty)$ which is very hard to be accepted for the analysis of maximal hydrological parameters, and the asymmetrical coefficient is taken as constant value ($C_s \approx 1,14$) because the distribution is two-parametric. The most often reason for application of Gumbel distribution for analysing the maximal hydrological values is the fact that the Gumbel distribution function values can never be smaller than 0, which also corresponds to the real hydrological phenomena's;

(3) Theoretical gamma distribution functions have wide application in mean and maximal hydrological parameter analysis, and in minimal as well. There are a numerous distribution functions of the gamma type, from which the broader used are: Kick and Menkeley distribution functions, Raileigh Weibull, and Erlang distribution functions, three-parametric Weibull distribution, and the most universal and widespread method in hydrological analysis all over the world, the Pirson III distribution function. This function gives best results in mean hydrological parameter analysis. However, the defective of this method is that when the ration of the asymmetrical and variance function is smaller than 2, ($C_s/C_v < 2$), the definition zone of this function gets enlarged gaining negative values as a result, which is in contradiction with the hydrological phenomena's in reality. However, in practice, in the case when the ratio $C_s/C_v < 2$, the value C_v is taken as a constant while the coefficient of asymmetry C_s takes various values until a satisfactory level of approximation is met between the theoretical function curve and the distribution frequency curve of the factual measurements. The Pirson III three-parametric distribution function of the type of the American standards, is an obligatory for comparison of maximal hydrological parameters obtained with other distribution functions

Considering the water flow characteristics measured at the Nadakoc station of the Sitnica river, we have developed one practical program, in excel,

for achieving probabilistic water-flow forecast for different time frequencies, using different theoretical distribution functions as normal distribution function (Gauss), log-normal, Gumbel, and Pirson III.

In Table 5, the some distribution functions taken from different authors are presented, as well as their corresponding probability distribution functions.

A very important step is the comparison of the true measured values with the theoretical one obtained from the theoretical distribution function. Therefore, in the following, we show in tabular form the density distribution function taken from four different authors and the factual data as well (Table 6). The same is also shown in graphical form in Fig. 5. This graph

shows the normalised values of distribution function and the factual data. Considering Fig. 5, we can conclude which distribution function corresponds best to the factual data taken from Nadakoc station.

In the following, we show in tabular and graphical form, the probability (cumulative) distribution function taken from the factual data, Gauss normal distribution, log-normal distribution, Gumbel exponential distribution and Pirson III gamma distribution (Fig. 6 and Table 7).

From Fig. 6 and Table 7, we see that the probability distribution of the different functions coincides more or less with the factic measured data.

The ultimate step would be the representation of

Table 5 Various distribution functions, and their corresponding probability distribution function.

Theoretical distribution function	Density distribution function (density curve)	Cumulative probabilistic distribution function (probabilistic curve)
Normal distribution function (Gauss)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{[-(x_i - \bar{x})^2 / 2\sigma^2]}$	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{[-(x_i - \bar{x})^2 / 2\sigma^2]} dx$
Two-parametric log-normal distribution	$f(x) = \frac{1}{x \cdot \sigma_y \sqrt{2\pi}} \cdot e^{[-(\ln x_i - \bar{x}_y)^2 / 2\sigma_y^2]}$	$F(x) = \frac{1}{x \cdot \sigma_y \sqrt{2\pi}} \cdot \int_{-\infty}^x e^{[-(\ln x_i - \bar{x}_y)^2 / 2\sigma_y^2]} dx$
Gumbel distribution function	$f(x) = e^{-z-e^{-z}}; \quad z = \alpha(x - M_0)$	$F(x) = 1 - e^{-e^{-z}}; \quad z = \alpha(x - M_0)$
Pirsonit III distribution function	$f(x) = c^{-\frac{x}{d}} \left(1 + \frac{x}{a}\right)^{\frac{a}{d}};$ $d = \frac{C_s \cdot C_v}{2}; \quad a = C_v \left(\frac{2}{C_s} - \frac{C_v}{2} \right)$	$F(x) = \int_{-a}^x y_0 \cdot e^{-\frac{x}{d}} \left(1 + \frac{x}{a}\right)^{\frac{a}{d}} dx$

Table 6 Density curve comparison table.

Factual occurrence ratio (density curve)	Gauss	Log-normal	Gumbel	PirsonIII
18.52	10.69	1.73	12.79	8.11
3.70	15.34	11.33	20.50	17.86
22.22	19.24	23.36	23.07	22.68
14.81	21.09	25.71	20.82	22.78
18.52	20.21	19.07	16.37	20.34
7.41	16.93	10.78	11.80	17.04
3.70	12.40	4.99	8.06	13.76
3.70	7.94	1.99	5.33	10.88
3.70	4.44	0.70	3.45	8.50
0.00	2.17	0.23	2.20	6.60
0.00	0.93	0.07	1.39	5.11
3.70	0.35	0.02	0.88	3.96

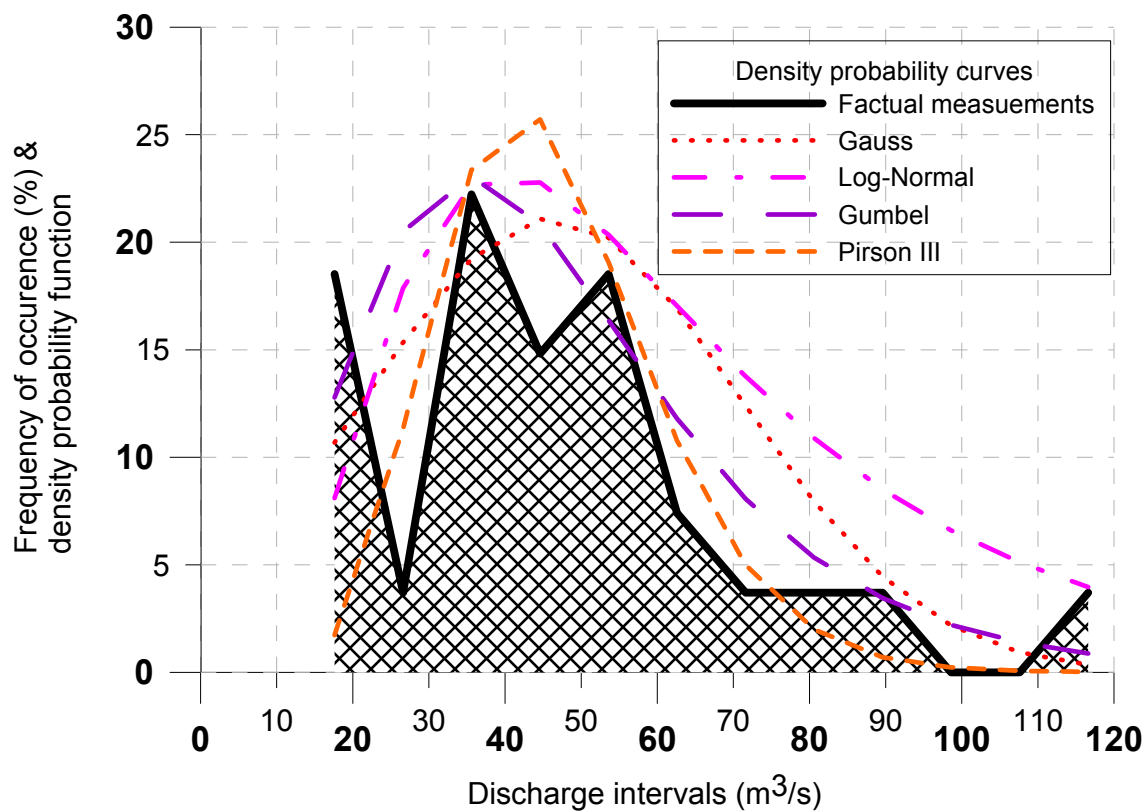


Fig. 5 Normalized distribution function curves.

Table 7 Probability curve comparison table.

Factual measurements	Gauss	Log-normal	Gumbel	Pirson III
18.52	12.14	7.36	8.12	0.49
22.22	21.15	23.50	20.83	6.28
44.44	33.20	41.72	37.53	24.06
59.26	47.31	57.45	54.21	49.58
77.78	61.77	69.55	68.22	72.41
85.19	74.74	78.38	78.75	87.22
88.89	84.92	84.65	86.13	94.85
92.59	91.92	89.07	91.10	98.15
96.30	96.14	92.17	94.34	99.39
96.30	98.36	94.35	96.43	99.82
96.30	99.38	95.90	97.75	99.95
100.00	99.79	97.00	98.59	99.99

maximum river-flow (discharge) as a function of return period of occurrence, in a semilogarithmic graphical format (Fig. 7).

From Fig. 7, it can be seen that for different return period of time, different maximal water-flows (discharge) are expected, but always depending on the theoretical distribution function. It is essential to be able to choose the most appropriate distribution

function which would best fit the measured data. Therefore, in evaluating the hydrological parameters (as maximal water-flow), it is most necessary to integrate different distribution functions in our analysis.

From Fig. 7, it can be ultimately concluded that, the best theoretical probability function which would best fit the factual data taken from a 30 year period in the

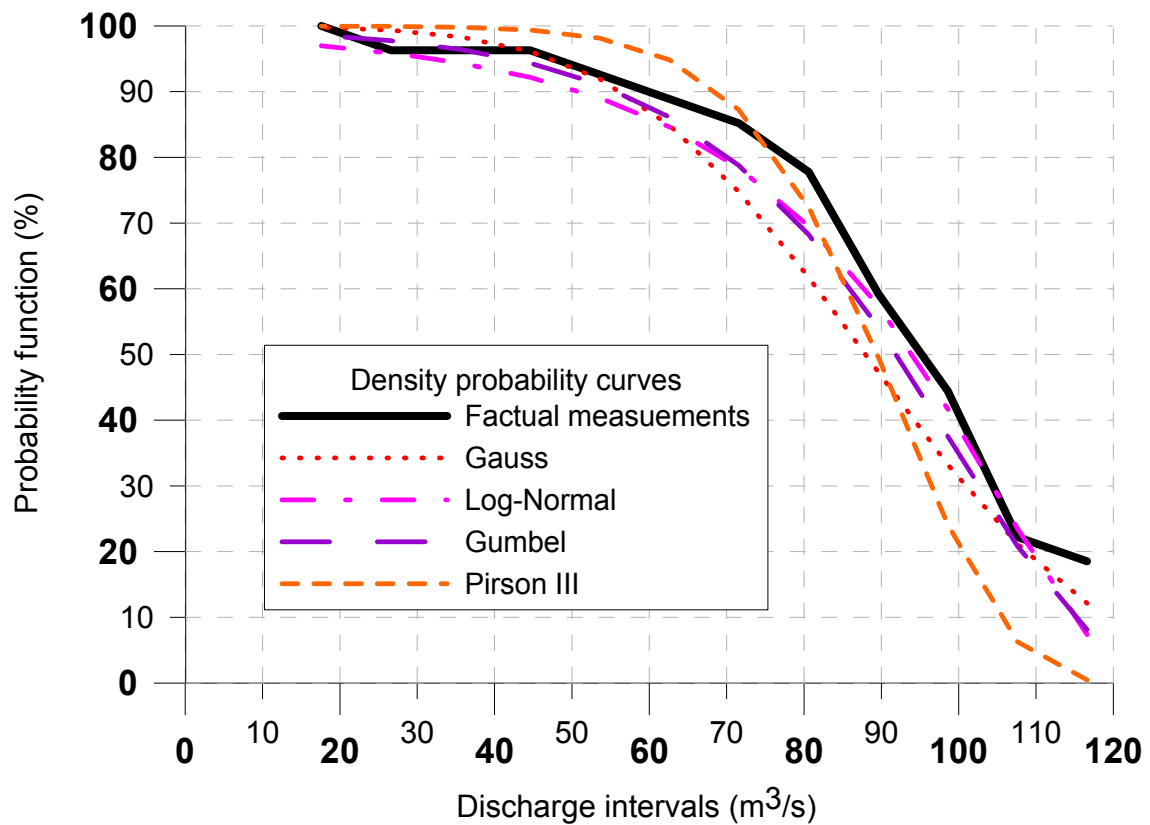


Fig. 6 Probability curves.

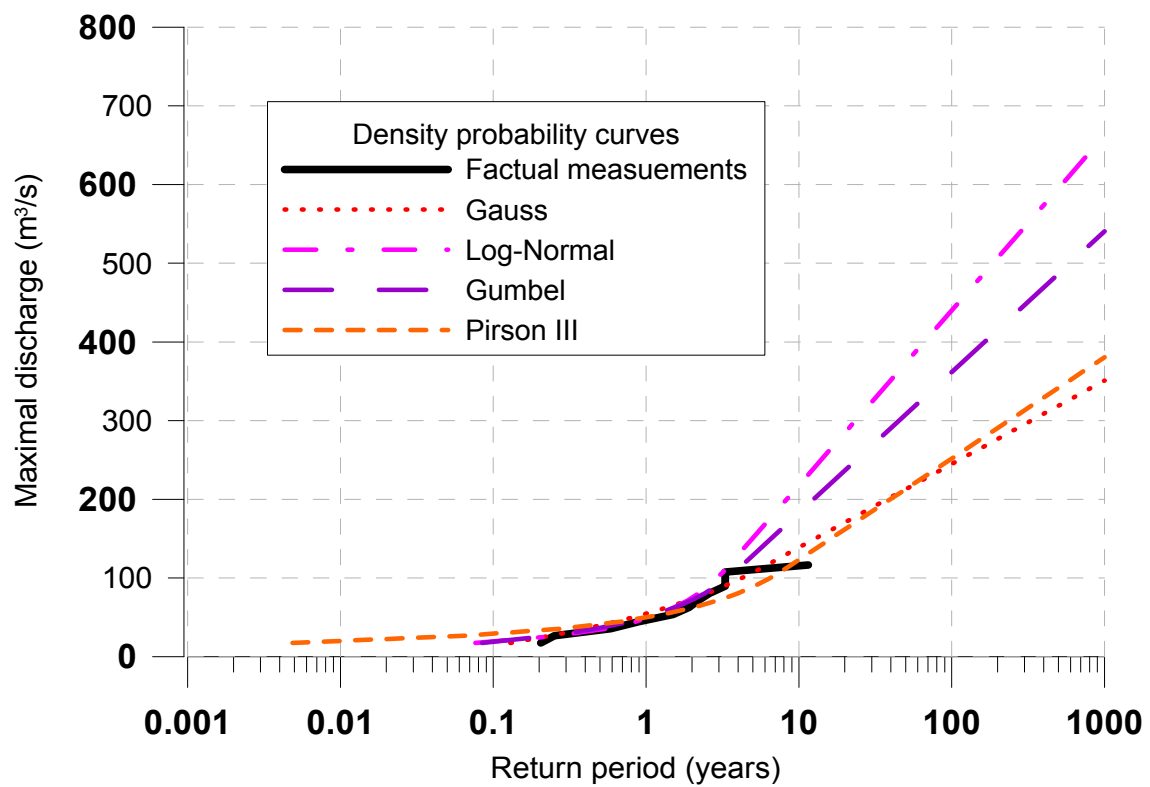


Fig. 7 Maximal discharge graph as a function of return period.

Table 8 Anticipated maximal river discharge for the Sitnica river.

Return period	Maximal anticipated river discharge (m ³ /s)
1	40
10	110
100	250
200	290
500	350

Nadakoc station of the Sitnica river, would be the Gaussian normal distribution function as well as Pearson III gamma distribution function.

Considering the Pearson III distribution function, in Table 8, the anticipated maximal river discharge for specific return periods is shown.

Further on, these very important data are used in the further analysis and design in hydrological problems.

As a conclusion, it could be emphasised once more the importance of probabilistic approach in hydrology which is the essential principal in various hydrological analysis. In the modern hydrological analysis, can be clearly seen the vastness of applicability of theoretical

distribution functions in hydrology, especially Gauss normal distribution function, Log-normal two-parametric distribution function, Gumbel exponential distribution function and Pearson 3 gamma distribution function. These four theoretical functions today have become standard tools for statistical analysis in hydrology.

References

- [1] Shehu, B., Karanxha, K. 2003. *Hidrologjia Inxhinierike 1*. Tiranë: Shtëpia Botuese e Librit Universitar (SHBLU). (in Albanian)
- [2] Schervish, D. 2011. *Probability and Statistics*. 4th ed. New York: Pearson.
- [3] Soong, T. T. 2004. *Fundamentals of Probability and Statistics for Engineers*. 4th ed. New York: John Wiley and Sons.