

Modelling the annual maxima discharges of the Jihlava river applying Weibull probability distribution

Mária Michalková¹
Zuzana Sedliacková²
Ivana Pobočíková³
Martina Kozáčková⁴

¹ Department of Applied Mathematics, Faculty of Mechanical Engineering, University of Žilina; Univerzitná 8215/6, 010 15 Žilina, Slovakia, maria.michalkova@fstroj.uniza.sk

² Department of Applied Mathematics, Faculty of Mechanical Engineering, University of Žilina; Univerzitná 8215/6, 010 15 Žilina, Slovakia, zuzana.sedliackova@fstroj.uniza.sk

³ Department of Applied Mathematics, Faculty of Mechanical Engineering, University of Žilina; Univerzitná 8215/6, 010 15 Žilina, Slovakia, ivana.pobocikova@fstroj.uniza.sk

⁴ Department of Applied Mathematics, Faculty of Mechanical Engineering, University of Žilina; Univerzitná 8215/6, 010 15 Žilina, Slovakia, martina.kozackova@fstroj.uniza.sk

Grant: KEGA 031ŽU-4/2026

Name of the Grant: **Innovative approaches to teaching Mathematics 1 for future engineers: The concept of Blended Learning**

Subject: BB – applied statistics, operational research

© GRANT Journal, MAGNANIMITAS Assn.

Abstract Reliable prediction of extreme hydrological events, such as floods, requires two key elements – a sufficiently long continuous series of observations and selection of an appropriate probability distribution. Here, we analyze a 98-year long continuous series of average daily discharge records of the Jihlava River at the Dvorce gauging station. The dataset of annual block maxima is modeled by a two-parameter Weibull distribution. The distribution parameters are estimated using the maximum likelihood method. The Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests are employed to verify the agreement between the theoretical model and the observed data. Based on the p-values of both tests, which exceed the 0.05 significance level, we can conclude that the selected probability distribution adequately models the annual maximum of average daily discharges in the analyzed section of the Jihlava River.

Keywords block maxima, Weibull probability distribution, gauging station Dvorce

1. INTRODUCTION

Systematic monitoring of daily river discharge holds strategic importance for safety, the economy, and environmental protection. Long-term records enable more accurate modeling for designing infrastructure such as bridges, dams, and levees. Monitoring is also crucial for early warning systems against extreme hydrological events like floods or the onset of drought, providing the time needed to minimize their impact. Effective risk management relies on a dependable probabilistic model. For flood management, such a model serves as the foundation for estimating the magnitude and return period of extreme flows, their frequency over time, and the probability of exceeding specific extreme thresholds.

The range of probability distributions applicable to modeling extreme events is quite limited. Commonly used options include the Weibull, Gumbel, and Fréchet distributions (collectively known as the generalized extreme value distribution), the Pearson type III and log-Pearson type III distributions, and the generalized Pareto distribution. In [1], maximum annual discharges were modeled at seven hydrological stations in India's Brahmani and Baitarani river basins. The authors compared the goodness of fit between measured flows and four probability distributions: the generalized extreme value, Gumbel, log-Pearson type III, and log-normal distributions. They estimated parameters using the maximum likelihood, moments, and L-moments methods. In [2], the authors compared parametric and non-parametric models for the highest annual peak flows across 18 stations in the Punjab region of Pakistan, evaluating how well empirical observations fit the generalized Pareto, generalized extreme value, and generalized logistic distributions. Karim et al. [3] applied five probability distributions to data from 24 gauging stations in the Great Barrier Reef lagoon catchments, while also comparing frequency estimates derived from the block maxima and peak-over-threshold methods. Coufal and Hornová [4] evaluated N-year discharges at three gauging stations (Raškov, Kroměříž, and Strážnice) in the Morava River basin, approximating the maximum annual flow series with four distributions (log-normal, Pearson type III, generalized extreme value, and Gumbel) and comparing their accuracy using mean squared error.

Our objective is to investigate the applicability of the 2-parameter Weibull distribution for modeling the annual maxima of average daily discharges on the upper reaches of the Jihlava River at the Dvorce hydrological station. The distribution parameters are estimated via the maximum likelihood method using MATLAB R2020b software. Finally, the fit of the Weibull distribution to the empirical data is verified using the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests.

2. STATISTICAL MODEL

There are two main approaches to selecting maximum values from a dataset: the block maxima method and the peak-over-threshold method.

The block maxima method divides a time-series dataset into equal, non-overlapping intervals, or blocks (such as a day, month, or year). Extracting the highest value from each block creates the set of block maxima. Choosing sufficiently long blocks helps ensure the observations are not correlated—which is important since shorter blocks often capture seasonal patterns common in hydrological data. However, a key disadvantage is that only one maximum is chosen per block. This can lead to ignoring significant second or third highest values in a specific block, even if they are much larger than the maximum of a different block.

Alternatively, the peak-over-threshold method sets a specific baseline, and any observation exceeding it is included in the dataset. This prevents the data loss associated with the block maxima method. The main drawback here is that these observations can be correlated, such as when a single flood wave spans several consecutive days. Selecting the right threshold value is also a challenge. [3, 5, 6]

The chosen selection method directly dictates which probability distributions apply. According to the Fisher-Tippett-Gnedenko theorem [7, 8], the probability distribution of block maxima converges to one of three distributions:

- Gumbel distribution,
- Fréchet distribution,
- Weibull distribution.

2.1 2-parameter Weibull distribution

The probability density function $g(x)$ and the cumulative distribution function $G(x)$ of the 3-parameter Weibull distribution are defined, respectively,

$$g(x) = \frac{a}{b^a} x^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a\right)$$

and

$$G(x) = 1 - \exp\left(-\left(\frac{x}{b}\right)^a\right)$$

for $x > c, a > 0, b > 0$. Parameter a is the shape parameter, parameter b is the scale parameter and parameter c is the location parameter.

The maximum likelihood estimates of the parameters are obtained by solving the system of equations

$$\frac{1}{a} - \frac{\sum_{i=1}^n x_i^a \ln x_i}{\sum_{i=1}^n x_i^a} + \frac{1}{n} \sum_{i=1}^n \ln x_i = 0,$$

$$b = \left(\frac{1}{n} \sum_{i=1}^n x_i^a \right)^{\frac{1}{a}}.$$

2.2 Goodness-of-fit tests

To verify agreement between the theoretical and empirical distribution, two goodness-of-fit tests are used: The Kolmogorov-Smirnov test and the Anderson-Darling test. Let $\hat{F}(x)$ be the estimated distribution function, and $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be

a realization of a random sample ordered in ascending size, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.

We test the null hypothesis:

H_0 : the random sample originates from the theoretical distribution against the alternative

H_1 : the random sample does not originate from the theoretical distribution.

The test statistic for the Kolmogorov-Smirnov test is given by

$$D = \max_{1 \leq i \leq n} \left[\left| \hat{F}(x_{(i)}) - \frac{i-1}{n} \right|, \left| \frac{i}{n} - \hat{F}(x_{(i)}) \right| \right].$$

The test statistic for the Anderson-Darling test is given by

$$A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} \left[\ln(\hat{F}(x_{(i)})) + \ln(1 - \hat{F}(x_{(n+1-i)})) \right].$$

We use the p -value to test agreement. Hypothesis H_0 is rejected at significance level α if $p\text{-value} < \alpha$.

3. DESCRIPTION OF STUDIED LOCATION

The Jihlava River originates in the Bohemian-Moravian Highlands in hilly terrain near the village of Jihlávka. The total length of the river is 184.5 km; the basin area is 2998 km². Together with the Svatka River, it empties into the Nové Mlýny middle reservoir. The largest tributaries of the Jihlava River are the Oslava (left-bank tributary) and Rokytá (right-bank tributary) rivers. Major towns located along the course of the river are Jihlava, Třebíč, Ivančice, Dolní Kounice, and Pohořelice. Two reservoirs were constructed on the river—Dalešice and Mohelno—which also serve as a water supply for the Dukovany Nuclear Power Plant. [9, 10]



Fig. 1. Hydrological station 465000 Dvorce – map location (upper); station detail (lower). [11]

Hydrological gauging station 465000 Dvorce (49° 22' 50.18", 15° 30' 15.38") is located near the village of the same name. The station lies at river kilometer 150.64, i.e., approximately 34 kilometers from the river source, at an elevation of 501.17 m a.s.l. (gauge datum zero). The station itself is situated on the right bank of the river, 50 m below the bridge at an old dead-end road (Fig. 1). The station operator is Povodí Moravy Brno. The first records from this station date back to 1928. Currently, water level is recorded at the station, and water discharge is calculated from the water level. The average annual water level is 39 cm, corresponding to a discharge of 1.77 m³/s. Over the course of measurements, several floods were recorded in this reach. The highest water level of 272 cm (discharge 53.40 m³/s) was measured on April 1, 2006, during a flood that affected most of the Czech Republic. [11, 12]

3.1 Data characteristics

Continuous records of average daily discharges exist for the Dvorce station from the beginning of measurements in 1928 up to and including 2025. Thus, the dataset used [13, 14] contains 35,795 records (from January 1, 1928, to December 31, 2025). From these data, a set of maximum annual discharges comprising 98 values was created using the block maxima method. The block size was one calendar year (daily averages from January 1 to December 31 of each calendar year). Descriptive statistics for daily averages as well as annual maxima are summarized in Table 1.

The lowest average daily discharge of 0.02 m³/s was recorded at the station on October 31, 1951. On the preceding two days, the second lowest average discharge in measurement history was recorded at 0.04 m³/s. Conversely, the largest average daily discharges occurred during devastating floods at the turn of March and April 2006 (from March 30 to April 1, average daily discharges reached 38.70 m³/s, 41.50 m³/s, and up to 45.10 m³/s). The third largest average daily discharge was recorded on March 20, 1947, at 39.50 m³/s, caused by an ice flood. As shown in Fig. 3, the year 2025 can be considered hydrologically the weakest, when the maximum of all average daily discharges reached only 2.26 m³/s. Similarly, dry years were 2018 (maximum average daily discharge was 3.04 m³/s) and 2016 (maximum average daily discharge was 3.97 m³/s).

The mean value of annual maxima is 15.547 m³/s. The values exhibit a high degree of variability, which is confirmed by a coefficient of variation of 56.54%. The skewness coefficient for annual maxima indicates a positively skewed probability distribution (Fig. 2); the kurtosis coefficient indicates a leptokurtic distribution.

Table 1. Descriptive statistics of the dataset of average daily discharges and annual maximum discharges.

	Average Daily Discharge	Annual Maxima of Average Daily Discharge
Mean (m ³ /s)	1.873	15.547
Standard Deviation (m ³ /s)	2.192	8.791
Lower Quartile (m ³ /s)	0.80	9.49
Median (m ³ /s)	1.26	13.50
Upper Quartile (m ³ /s)	2.10	19.40
Kurtosis	52.991	1.015
Skewness	5.588	1.088
Minimum (m ³ /s)	0.02	2.26
Maximum (m ³ /s)	45.10	45.10
Coefficient of Variation (%)	117.05	56.54
Number of Measurements / Blocks	35 795	98

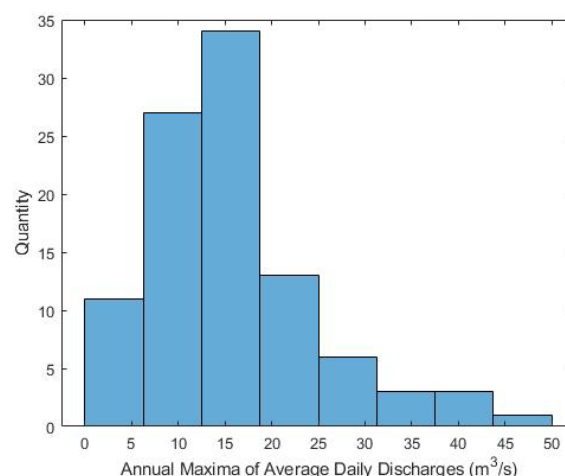


Fig. 2. Histogram of annual maxima of average daily discharges.

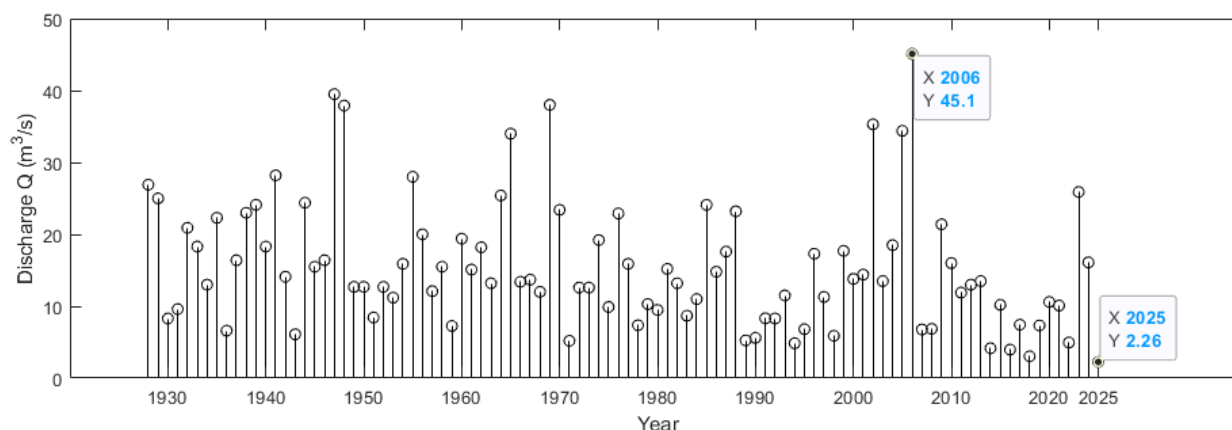


Fig. 3. Annual maxima of average daily discharges at the Dvorce station.

4. RESULTS

Parameter estimation for the 2-parameter Weibull distribution is summarized in Table 2.

Tab. 2. Parameter estimates.

shape parameter $\hat{a}$	1.895
scale parameter $\hat{b}$	17.601

Results for the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests are presented in Table 3.

Tab. 3 Goodness-of-fit tests.

	Test Statistic	p-value
Kolmogorov-Smirnov test	0.080	0.526
Anderson-Darling test	0.092	0.648

The p-value for both tests is greater than the chosen significance level $\alpha = 0.05$, which means we do not reject hypothesis H_0 . Block maxima of average daily discharges on the Jihlava River follow a Weibull distribution, as confirmed by the histogram overlaid with the distribution curve (Fig. 4) as well as the Q-Q plot (Fig. 5).

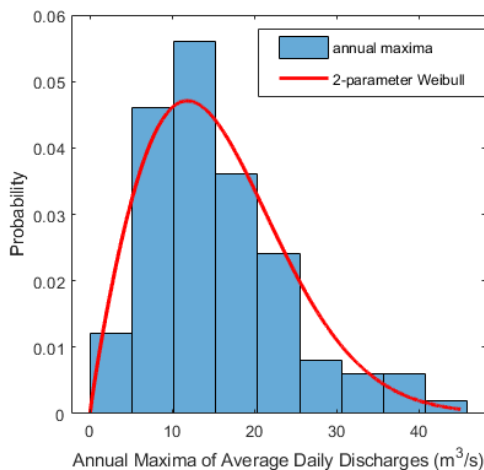


Fig. 4. Histogram for annual maxima of average daily discharge overlaid with the estimated Weibull distribution curve.

5. RETURN PERIOD AND RETURN LEVEL

The return level Q_N is the value that is exceeded on average once in N years with the probability [15, 16]

$$p = \frac{1}{N}$$

The probability that the discharge exceeds the return level is then defined as

$$P(Q > Q_N) = 1 - G(Q_N) = p$$

where $G(x)$ denotes the cumulative distribution function. Thus, the return level represents the quantile of the cumulative distribution function

$$Q_N = G^{-1}\left(1 - \frac{1}{N}\right).$$

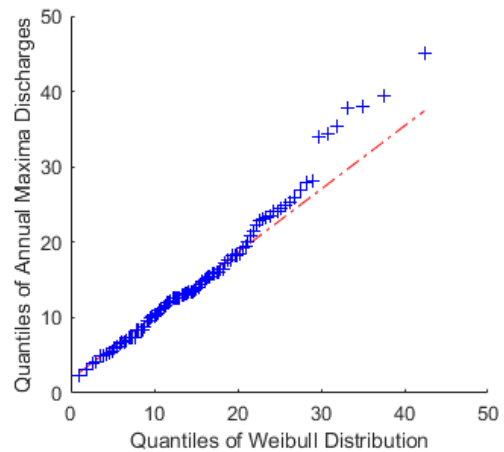


Fig. 5. Q-Q plot - quantiles of maxima discharges (blue cross), quantiles of Weibull distribution (red line).

For the 2-parameter Weibull distribution it is calculated as [16]

$$Q_N(N) = b \left(-\ln \left(1 - \frac{1}{N} \right) \right)^{\frac{1}{a}}.$$

The return period, N , is typically set at 2, 5, 10, 20, 50 or 100 years to determine N -year flood flows. For the estimation of the N -year discharge to be reliable, a time series of the measurements with a length of approximately $5N$ is required. [15]

In Table 4, we provide N -year discharges for $N = 5, 10, 20, 50$ and 100 years. Since we have 98 annual maxima available, estimation up to 20 years can be considered reliable, whereas higher return periods are less reliable. For example, in a 100-year long time series, the probability of Q_{100} occurring at least once is 63%. [15]

Table 4. Estimates of the N -year discharges at the hydrological station Dvorce.

N	5	10	20	50	100
Q_N (m³/s)	22.62	27.34	31.42	36.17	39.43

6. CONCLUSION

In this paper, we modeled extreme discharges on the Jihlava River using the block maxima method. Nearly a century of continuous average daily discharge records from the Dvorce hydrological station were used for the analysis. By choosing a block length of one year, 98 values — representing maximum average daily discharges for the years 1928 through 2025 — were extracted. A 2-parameter Weibull probability distribution was used to model the dataset. Distribution parameters were estimated using the maximum likelihood method (shape parameter $a = 1.895$, scale parameter $b = 17.601$). Agreement between the theoretical distribution and empirical data was confirmed by two goodness-of-fit tests. The p-value of the Kolmogorov-Smirnov test (0.526) and the Anderson-Darling test (0.648) consistently demonstrated that H_0 cannot be rejected at significance level $\alpha = 0.05$. Adequacy of the estimated Weibull distribution for block maxima of average daily discharges was also confirmed by the Q-Q plot and the distribution curve fitted over the histogram of observations.

Acknowledgments

This research was supported by the Cultural and Educational Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic (KEGA) through project No. 031ŽU-4/2026 "Innovative approaches to teaching Mathematics 1 for future engineers: Blended Learning concept".

References

1. KHWAIRAKPAM, R.S., KUNDU, S. Significance of different probability distributions in flood frequency analysis of Brahmani-Baitarani River Basin, India. *Discover Geoscience*, 2(71), 2024. 20s. doi 10.1007/s44288-024-00072-8
2. FAWAD, M., CASSALHO, F., REN, J., CHEN, L., YAN, T. State-of-the-Art Statistical Approaches for Estimating Flood Events. *Entropy*, 24, 2022. p. 898. doi: 10.3390/e24070898
3. KARIM, F., HASAN, M., MARVANEK, S. Evaluating annual maximum and partial duration series for estimating frequency of small magnitude floods. *Water*, 9, 2017. doi: 10.3390/w9070481
4. COUFAL, P., HORNOVÁ, H. Opakování velkých vod v povodí Moravy. *Hydrological Processes, Ecosystems and Climate Change* (ed. Vitková, J., Botyanszká, L.), Bratislava: Slovenská akadémia vied, Hydrologický ústav, 2024. 192 p. ISBN 978-80-89139-61-3
5. COLES, S. *An Introduction to Statistical Modelling of Extreme Values*. Londýn: Springer, 2001. 223 p. ISBN 978-1-85233-459-8
6. PAN, X., RAHMAN, A., HADDAD, K., OUARDA, T.B.M.J. Peaks-over-threshold model in flood frequency analysis: a scoping review. *Stochastic Environmental Research and Risk Assessment*, 36, 2022. pp. 2419-2435. doi: 10.1007/s00477-022-02174-6
7. FISHER, R.A., TIPPET, L.H.C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 1928, pp. 180-190. doi:10.1017/S0305004100015681
8. EMBRECHTS, P., KLÜPPELBERG, C., MIKOSCH, T. *Modelling extremal events (for insurance and finance)*. Berlin: Springer-Verlag, 1997. 648 p. ISBN 978-3-540-60931-5
9. POVODÍ MORAVY, *Významné řeky*. Online [cit. 30.6.2026] <https://www.pmo.cz/cz/uzitecne/vyznamne-vodni-toky/>
10. Jihlava. *Encyklopaedia Beliana*. ISBN 978-80-89524-30-3. Online [cit. 30.6.2026] <https://beliana.sav.sk/heslo/jihlava-0>
11. ČESKÝ HYDROMETEOROLOGICKÝ ÚSTAV, *Data ze stanice povrchových vod Dvorce*. Online [24.6.2026] <https://www.chmi.cz/namerena-data/merici-stanice/povrchove-vody/2505289-dvorce?c=49.3806,15.5043,14.0000&l=ORTHO#zahlaví>
12. ČESKÝ HYDROMETEOROLOGICKÝ ÚSTAV, *Evidenční list hlášeného profilu č. 385*. Online [24.6.2026] <https://hydro.chmi.cz/hppsevlst/download?seq=2505289&lng=CZE>
13. ČESKÝ NÁRODNÍ VÝBOR PRO HYDROLOGII, *Vybrané hydrologické a klimatologické datové sady ke stažení*. Online [cit. 19.6.2026] http://cnvh.cz/index.php?option=com_content&view=article&id=170&Itemid
14. ČESKÝ HYDROMETEOROLOGICKÝ ÚSTAV, *Vodoměrná stanice: 465000 - Dvorce. ISVS - Evidence množství povrchových vod*. Online [cit. 19.6.2026] https://isvs.chmi.cz/ords/f?p=11002:2:116126348427583:::RP,2:P2_SEQ:%5C459%5C
15. SOLÍN, Ľ. Povodne - odhad ich veľkostí pre malé povodia Slovenska metódou regionálnej frekvenčnej analýzy. *Geografický časopis*, 57(4), 2005. pp 287 - 307. ISSN 2453-8787 (online)
16. DE SOUZA, A., DE MEDEIROS, E. S., DE OLIVEIRA-JÚNIOR, J.F., CARVALHO ABREU, M., KUMAR, V., PESSOA BEZERRA, A. Assessing wind energy potential in southeast Brazil: fitting Weibull, Gumbel, and GEV distributions to wind speed data. *Geoenvironmental Disasters*, 12(34), 2025. pp. 12 - 34. doi: 10.1186/s40677-025-00337-5