



Article Comparison of Three-Parameter Distributions in Controlled Catchments for a Stationary and Non-Stationary Data Series

Łukasz Gruss^{1,*}, Mirosław Wiatkowski¹, Paweł Tomczyk¹, Jaroslav Pollert² and Jaroslav Pollert, Sr.²

- ¹ Institute of Environmental Engineering, Wrocław University of Environmental and Life Sciences, 50-363 Wrocław, Poland; miroslaw.wiatkowski@upwr.edu.pl (M.W.); pawel.tomczyk@upwr.edu.pl (P.T.)
- ² Faculty of Civil Engineering, Czech Technical University in Prague, 16629 Prague, Czech Republic; Jaroslav.Pollert@cvut.cz (J.P.); Jaroslav.Pollert.2@cvut.cz (J.P.S.)
- * Correspondence: lukasz.gruss@upwr.edu.pl; Tel.: +48-71-320-5518

Abstract: Flood Frequency Analysis (FFA) and the non-stationary FFA approaches are used in flood study, water resource planning, and the design of hydraulic structures. However, there is still a need to develop these methods and to find new procedures that can be used in estimating simple distributions in controlled catchments. The aim of the study is a comparison of three-parameter distributions in controlled catchments for stationary and non-stationary data series and further to develop the procedure of the estimation the simple distributions. Ten rivers from the Czech Republic and Poland were selected because of their existing or planned reservoirs as well as for flood protection reasons. The annual maximum method and the three-parameter Weibull, Log-Normal, Generalized extreme value, and Pearson Type III distributions were used in this study. The analyzed time series are stationary and non-stationary. The methodology used in this study, which makes use of the Maximum Likelihood Estimation, allows one to simplify the analysis whenever there is a series of data that is both stationary and non-stationary. The novelty in our research is the standardization and development of a new procedure for a stationary and non-stationary data series, taking into account to read a specific value of the maximum flow with a given exceedance probability from the lower or upper tail. It determines the optimal choice of the theoretical distribution that can be used, for example in the design of weirs in rural areas (lower quantiles) or in the design of hydrotechnical structures in areas at risk of flooding (upper quantiles).

Keywords: flood frequency analysis; non-stationary; three-parameter distribution; GEV distribution; Weibull distribution; Log-Normal distribution; Pearson Type III distribution; maximum likelihood

1. Introduction

In many parts of the world, extreme natural phenomena such as extreme streamflows and floods lead to loss of life and property [1,2]. They cause destruction of infrastructure [3–5] and damage the environment and economy [4,5].

In several studies [2,6,7], an appropriate estimation of flow values for various probabilities of occurrence or various return periods is required. Examples include the design of hydrotechnical structures and water management [4,8–14]. Numerous methods are available for flood analysis using historical data [4,5]. Such methods include the at-side FFA, which allows one to select the most appropriate probability distribution and the estimation of its adequate parameters [4,5,11,15]. such as climate and landcover, are the same in the past, present, and future [16]. Otherwise, the non-stationary FFA approach is recommended [10,16–18]. The relationship between the flow and the probabilities of occurrence is unique for every gauging station [14,18–20]. The maximum flow peak value occurring in a water year at a given cross-section of a stream collected for the longest possible period of observation allows one to get more realistic results of the FFA [1,13,21]. Moreover, it is very important because the obtained estimated values of the parameters of the probability



Citation: Gruss, Ł.; Wiatkowski, M.; Tomczyk, P.; Pollert, J.; Pollert, J., Sr. Comparison of Three-Parameter Distributions in Controlled Catchments for a Stationary and Non-Stationary Data Series. *Water* 2022, *14*, 293. https://doi.org/ 10.3390/w14030293

Academic Editor: Mojca Šraj

Received: 10 December 2021 Accepted: 15 January 2022 Published: 19 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). distribution function should be unbiased and close to their population values [1]. The maximum peak flow of each year in the longest observation period consists of the sample series and is used in the annual maximum (AM) approach [4,5,20–25]. The AM is still the most commonly used method in at-side FFA [12,20,24,26,27] and in non-stationary FFA in many countries [10,16–18], including the Czech Republic and Poland.

In the hydrological literature [5,21,26–28], traditional two-parameter distributions are most commonly studied. Despite the satisfactory results yielded by these distributions, the use of multi-parameter distributions could lead to more appropriate fits between simulated and observed data [11,15]. Therefore, the use of multi-parameter distributions has been proposed, namely three-parameter distributions such as the Generalized Extreme Value (GEV), the three-parameter Log-Normal (LN3), the three-parameter Weibull, and the Pearson Type-III (3PIII). These distributions are recommended for an at-side FFA in many countries and are common in the literature [4,5]. Apart from the scale and shape parameters, a three-parameter distribution has a third parameter—the shift (location). This last parameter is added in order to reflect the long-term river flow regime and to increase the flexibility of models in terms of a variety of datasets [29]. These distributions have been studied by many scientists [5,11,13,18,30]. The selection of a fitting two- or three-parameter distribution usually depends on the characteristics of investigated data series at a particular site [4]. Therefore, it is very important to study the best fit of the different distributions to the observational flow series of data streams.

For the selected distribution, an analysis should be performed, including the estimation of its parameter values. One of the most theoretically sound and extensively used methods for fitting probability distributions to hydrological data is the Maximum Likelihood Estimation (MLE) [5,17,31]. The MLE has been used for the 3LN, 3PIII [5], GEV, 3PIII [4,32], GEV [13,21,22,33], as well as GEV and 3LN [19]. MLE estimators are usually asymptotically efficient [22,32]. The MLE can be used to estimate the statistical parameters of distributions by examining the non-stationary data series [16] and in long time series [18]. In turn, other methods of estimating distribution parameters, such as the L-moments or probability weighted moment, are not recommended in non-stationary FFA [30]. As shown in the literature review, despite the high potential, the use of the MLE to estimate threeparameter distributions for non-stationary flow data series is marginal and unpopular. However, this topic requires further research under various conditions.

Both in the territory of the Czech Republic (in the Vltava and Morava basins) and Poland (in the Upper Oder basin), studies of simple distributions have been few and far between. The research does not cover all the available stations in Poland, especially in the area of the Upper Odra basin [34]. Recently, in 2008, in the research area, studies of three-parameter Log-Normal and two-parameter distributions such as the Log-Normal, Gumbel, and Fechet distribution were used for the Vltava river, Prague station [26]. Later, only two-parameter distributions such as the Log-Normal, Log-Pearson III, Gumbel, and Fechet were tested. The latter two distributions did not seem to fit well with the empirical data [27]. Moreover, the Annual Maxima method and the GEV distribution were used for individual stations of the Morava river [33] and for the Elbe river, Litoměřice station [35]. The data series consisted of maximum annual water levels [33] and maximum annual flows [35]. For some of the rivers in the Upper Odra basin, including the Widawa in the Zbytowa profile, researchers used single distributions such as 2P2, 2LN, and GEV [13]. In turn, for the Bezimienny Potok (Minkovický Potok) located on the Polish-Czech border, researchers used Pearson Type III with the MLE [36]. In Poland, there were more studies of single distributions in the Vistula basin [12,18,19,30,37,38].

Ten rivers were selected because of their existing or planned reservoirs, as well as for flood protection reasons. The justification for this article is that the study of FFA and non-stationary FFA will allow one to analyze the possible distributions to be used for flood protection and engineering calculations in the design of water management facilities, which eventually should lead to improved water management. Flood Frequency Analysis (FFA) and the non-stationary FFA approaches are used in flood study, water resource planning, and the designing of hydraulic structures. However, there is still a need to develop these methods and to find new procedures that can be used in estimating simple distributions in controlled catchments.

The aim of the study is comparison of three-parameter distributions in controlled catchments for a stationary and non-stationary data series and further to develop the procedure of the estimation the simple distributions.

In addition, this study will, in the practice of flood protection, determine the optimal choice of theoretical distribution that can be used in the areas at risk of flooding—both urban and rural, or during the design of hydrotechnical structures.

2. Materials and Methods

2.1. Study Area

The maximum annual data series comes from 10 river gauging stations (Table 1) located in the Czech Republic and Poland (Figure 1). The source of data is the Czech Hydrometeorological Institute and the Institute of Meteorology and Water Management—National Research Institute. These data have been processed.

Table 1. Summary of the selected gauge stations.

No.	Country	River Name	Station	River Abbreviation	Station ID	Catchment Area [km ²]
1	Czech Republic	Berounka	Beroun	BB	198000	8286.23
2	Czech Republic	Labe	Děčín	LD	240000	51,120.34
3	Czech Republic	Morava	Olomouc-Nové Sady	МО	367000	3323.59
4	Czech Republic	Bečva	Teplice nad Bečvou	BT	389000	1275.32
5	Czech Republic	Morava	Kroměříž	MK	403000	7013.27
6	Poland	Bystrzyca	Krasków	ВК	150160120	683.40
7	Poland	Opawa	Branice	OB	150170160	604.46
8	Poland	Prudnik	Prudnik	PP	150170110	134.40
9	Poland	Ślęza	Białobrzezie	SB	150160250	181.00
10	Poland	Widawa	Zbytowa	WZ	151170050	720.7

Source: [39,40], own work.





Figure 1. Locations of analyzed gauging stations. Source: [39,40], own work.

The study area covers three river basins in these countries. The Berounka and Labe are located in the Labe basin, and the Morava and Bečva are located in the Morava basin. In turn, all rivers from Poland are located in the Upper Odra.

The selected rivers from the Labe and Morava basins have a much larger catchment area than the rivers from the Upper Odra basin (Table 1). The data of the annual maximum flow of ten gauging stations are examined in this investigation. The rivers from the Czech Republic also have a longer data series than those from Poland (Table 2). The length of the samples varies from 49 to 133 years (Table 2).

The rivers selected for the analysis, on the one hand, play an important role in the water management of both the Czech Republic and Poland, but on the other hand, they cause a flood risk. Reservoirs are planned on the rivers Labe, Morava, Becva, and on the tributaries of the Berounka River, on the Czech Republic side [41]. In turn, there are 12 existing reservoirs on the Morava tributaries, which are part of the water regulation works for flood protection. Moreover, the purpose of the reservoirs built in the 1960s–1970s is to supply both drinking water and service water [42]. Multifunctional reservoirs are planned on three of the five rivers (the Prudnik, Ślęza, Bystrzyca) near the gauging station, on the Polish side [43–46]. Lastly, on the Widawa River, there is a multifunctional reservoir [25]. The Opawa river was chosen because of the flood protection in its valley.

Two data collections, OB and WZ (N = 49) have the smallest sample size and the largest LD (N = 133) (Table 2). The data series of the WZ has a skewness of 0.30, which means it is fairly symmetrical. In turn, the data series of the SB has a skewness of 0.92, which means it is moderately skewed. Moreover, the remaining data series have the skewness above 1.0, which means they are highly skewed (Table 2). The BB, BK, BT, MK, MO, OB, and PP data series have a kurtosis greater than 3, which means they have heavier tails than the normal distribution. Moreover, the LD, SB and WZ data series have a kurtosis of less than 3, which means they have lighter tails than the normal distribution (Table 2).

River Abbreviation	Period	Ν	Mean [m ³ ·s ^{−1}]	Skewness [m ³ ·s ⁻¹]	Kurtosis [m ³ ·s ^{−1}]	Min [m ³ ·s ^{−1}]	Max [m ³ ·s ^{−1}]	$SD \\ [m^3 \cdot s^{-1}]$	CV [-]
BB	1911–2018	77	290	3.245	15.58	45.6	1680	235.6	0.8127
LD	1887–2018	133	1478	1.336	2.611	186	4600	743.8	0.5034
МО	1920–2018	98	173.6	2.143	8.116	52.7	686	94.84	0.5464
BT	1921–2019	92	194.7	2.536	9.805	29.9	841	119.3	0.6126
МК	1916–2018	103	356.8	1.315	3.473	119	1030	145.4	0.4075
ВК	1951–2019	69	69.53	2.103	4.197	8.06	371	77.34	1.112
OB	1967–2019	49	63.94	4.105	20.92	11	432	65.72	1.028
PP	1956–2019	64	39.73	2.161	4.964	1.95	220	45.71	1.151
SB	1951–2019	68	7.738	0.9182	-0.4071	0.33	22.2	6.083	0.7861
WZ	1971–2019	49	19.78	0.3011	-0.8346	4.63	39.8	9.701	0.4905

Table 2. Descriptive statistics.

Explanations to the table: Mean—Mean flow; SD—Standard Deviation of all values; CV—Coefficient of Variation of all the values [–].

If a data series has N < 90 (which applies to samples BB, BT, LD, MO, and MK), then the SD is greater. Otherwise, the SD is smaller (SD < 70 m³·s⁻¹ for BK, OB, PP, SB, and WZ) (Table 2). The highest CV was recorded for BK CV = 1.11 and the lowest was recorded for MK CV = 0.41 (Table 2).

The characteristics and the classification of the hydrological regime is presented in Appendix A (Table A1). The hydrological regime depends on the number of floods per year and the source of supply (Parde classification) and the variability of flows per year (Dynowska classification) [47–52]. The supplement in Appendix A (Figure A1) shows the variability of long-term monthly flow rates. Most rivers are dominated by the nival regime, while the remaining rivers are dominated by the pluvial-nival or nival-pluvial regime. This means that the spates caused by rainfall or thawing snow are responsible for the higher water flows in Poland and the Czech Republic (if the regime is simple—there was one clear spate in the year, if composite—there were two or more). At the same time, all rivers, except for the PP, are classified as lowland or upland. Seven out of 10 rivers are characterized by average balanced flows, which differs throughout the year (except for the BK, PP, and SB, where the variation during the year is low).

2.2. Change Point Analysis

Change Point Analysis is used to check the presence of abrupt changes in the mean of the annual peak in the data series records [53]. The FlowScreen R package was used to plot a metric with change points and trend [54,55]. The Kendall test was used to determine the trend. The change point is printed as solid black lines. If the temporal trend is significant (*p*-value < 0.05), the trend is plotted as a dotted red line for a trend [54].

2.3. Temporal Trend Analysis

The non-parametric Mann–Kendall trend test was used to detect the presence of temporal trends [11,56,57] in the annual maximum peak flow. The null hypothesis H0 is that there is no trend in the series, the alternative hypothesis H1 is that there is a trend [11,53]. Kendall's statistic S is calculated according to Equation (1):

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(x_j - x_k)$$
(1)

where x corresponds to the x univariate time series, n is the length of the maximum series, and j and k are time indices associated with individual values [11]. The S and

p-value were determined for each time series. In the Mann–Kendall trend test [11,20], the significance level was set at 5% [58]. The R Software version 4.0.5 was used to compute the Mann–Kendall trend test [55].

2.4. Test for Randomness

The rank version of von Neumann's ratio test (RVN) proposed by Bartels was used for testing a series for randomness [59,60]. The test statistic is given in Equation (2):

$$RVN = \frac{\sum_{i=1}^{n} (r_i - r_{i+1})^2}{\sum_{i=1}^{n} (r_i - r)^2}$$
(2)

where r_i is the rank of the ith observation in a sequence of n. The *p*-value is determined with the beta distribution over the range $0 \le x \le 4$ with parameters 3:

$$\alpha = \beta = \frac{5n(n+1)(n-1)^2}{2(n-2)(5n^2 - 2n - 9)} - \frac{1}{2}$$
(3)

where α and β are shape parameters of the distribution. The sample sizes should be in the range of $10 \le n < 100$. For large values of the sample size (n > 100), the approximation with N(2, 20 = (5n + 7)) is used to calculate the *p*-value. The null hypothesis H0 that the sample is random is tested against the alternative hypothesis H1 that the data are significantly different from random. In this test, the significance level was set at 5%. The DescTools R package was used to compute the rank version of von Neumann's ratio test.

2.5. PDF Fitting by the MLE

The MLE method estimates the parameters of a probability distribution by maximizing the likelihood function (4). The likelihood function is given by Equation (4) [5,28]:

$$\mathcal{L}(\theta) = f(x_1, x_2, \dots, x_n | \theta)$$
(4)

where $L(\theta)$ is called the likelihood function and *x*1, *x*2, and *x*3 are observed data.

If we have an empirical sample series, we can calculate the probability of occurrence of an observed sample series of any random variable by multiplying the PDFs of each empirical data of that series by each other, under the assumption that the events of the random variable are independent, which results in the likelihood function (LF) [1,5].

In order to find the Maximum Likelihood estimator, numerical optimization methods were used. For this purpose, the optim () function was used to find the ML estimator within the 3LN and 3W distributions [61] and the GEV distributions [62]. The nlminb () function was used to find the ML estimator within the 3PIII distribution [63]. The latter function was also used by Debele et al. [30]. The R Software version 4.0.5 was used to compute the MLE [55].

2.6. Probability Density Function (PDF)

The three-parameter probability distribution function of the 3LN was obtained by using the EnvStats R package and is given by the following formula [9,11,64]:

$$f(x) = \frac{1}{(x - \alpha)\sigma_{y\sqrt{2\pi}}} exp\{-\frac{1}{2\sigma_{y}^{2}}[log(x - \alpha) - \mu_{y}]^{2}\}$$
(5)

where μ_{y} , σ_{y^2} , α are the location, scale, and shape parameters, respectively.

The three-parameter Weibull distribution PDF expressed by Equation (6) is described by [65]:

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x-\mu}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x-\mu}{\beta}\right)^{\alpha}}$$
(6)

for $x > \mu$ and α , $\beta > 0$. The parameters α , β , and μ are the shape, scale, and shift parameters, respectively.

The PDF of the 3PIII distribution is given by:

$$f(x) = \frac{1}{|s|^{\alpha} \Gamma(\alpha)} |x - \lambda|^{\alpha - 1} e^{-\frac{x - \lambda}{s}}$$
(7)

for $s \neq 0$, a > 0, and $(x - \lambda)/s \ge 0$.

where α , *s*, and λ are the shape, scale, and location parameters, respectively.

The PDF function of the Generalized Extreme Value Distribution (GEV) is given in Equation (8) [11,13,20,21,66]:

$$f(x) = exp[-\left\{1 + \frac{s(x-\alpha)}{b}\right\}^{-1/s}]$$
(8)

for $1 + s(x - \alpha)/b > 0$, where b > 0; where α , b, and s are the location, scale, and shape parameters, respectively.

2.7. The Goodness-of-Fit Tests (GOF Tests)

The Goodness-of-Fit tests are used to check if the observed sample series follows an expected distribution.

The Anderson–Darling test has been applied to the comparison between the CDF from the tested distribution $F(x, \theta)$ and the empirical CDF Fn(x), where x is the studied variable, θ is a vector of parameters, and n is the number of elements in the sample. The null hypothesis is that H0:Fn(x) = $F(x, \theta)$, while the alternative hypothesis is that $F(x, \theta)$ is some other function. The significance level was set at 5% [4,11,20,32].

With the aid of the goftest R package, the Anderson–Darling test was performed to tests all of the observed AM series presented in relation with each sample series. The test makes use of Braun's method (1980) [67] to adjust for the effect of parameter estimation [68].

The following were used to define the best fit distribution [4,11,20]: the mean absolute error (MAE, Equation (9)) [4,11,20], the mean absolute relative error (MARE, Equation (10)) [13], and the root mean square error (RMSE, Equation (11)):

$$MAE(y) = \frac{1}{2} \sum_{i=1}^{n} \left| F(y_i) - F\left(\hat{y}_i\right) \right|$$
(9)

$$MARE(y) = \sum_{i=1}^{n} \left| \frac{y - y_i}{y_i} \right|$$
(10)

RMSE
$$(y) = \sqrt{\frac{\sum_{i=1}^{n} (F(y_i) - F(\hat{y}_i))^2}{n}}.$$
 (11)

Model estimation was accomplished by using the Akaike information criterion (AIC) [28] and the Bayesian information criterion (BIC) [5]. Before they can be calculated, both AIC and BIC require the likelihood to be maximized. The calculations are as follows [69]:

$$AIC = 2lnL(\theta) + 2k \tag{12}$$

$$BIC = 2lnL(\theta) + 2lnNk$$
(13)

where $L(\theta)$ is the value of the LF, *N* is the number of recorded measurements, and *k* is the number of estimated parameters.

The above-described GOF tests are often used in the FFA method [4,5].

The distribution is assigned a rank score between 1 and 4 in GOF tests. Rank score 4 indicates the best-fitted distribution, whereas 1 stands for the worst fitted distribution [4]. A distribution can obtain the highest score for the lowest value of the Anderson–Darling

statistic/highest *p*-value of the Anderson–Darling test. A distribution can also obtain the highest score for the lowest RMSE, lowest MARE, lowest MAE, lowest AIC, or lowest BIC. The total rank is the sum of all points in the row.

3. Results

Change points have been determined to be one of the possible causes of the nonstationarity of the annual flood peak time series [53]. We found (Figure 2a–j) that eight out of the 10 catchments exhibited a significant abrupt change in the mean. For BB, MO, and OB, the times of abrupt changes in the mean clustered around the 1990s and between 2000 and 2015 (Figure 2a,c,g). For LD, BK, PP, and WZ, the times of abrupt changes in the mean clustered between 2000 and 2015 (Figure 2b,f,h,j). Moreover, for BK and PP, the changes in the mean clustered around the 1950s and 1960s, respectively (Figure 2f,h). Only for MK and SB was no change point detected (Figure 2e,j).

On the one hand, the mean flow values for BB, LD, MO, BK, OB, PP, and WZ were much lower over the last period being analyzed after the change point (Figure 2a–c,f–h), but on the other hand, the mean flow values for BT (Q = 197.04 m³·s⁻¹ in 1921–1995, Q = 216.43 m³·s⁻¹ in 1995–2019) were much higher over the last period being analyzed after the change point (Figure 2d).

A second major cause of non-stationarity is a trend in the annual flood peak time series [53]. Analyses of the long-term trends were performed using the Mann–Kendall two-sided tests. A negative trend was detected only for LD and WZ. The Kendall test was also used to determine the trend. A temporal trend was detected for LD (*p*-value 0.008), WZ (*p*-value 0.01), but also for MK (*p*-value 0.04). This trend is decreasing (Figure 2b,c,e,j). The rank version of the von Neumann's ratio test proposed by Bartels did not show that the data were significantly different from random.



Figure 2. Cont.



Figure 2. Change points and trend for investigated data series: (**a**) Berounka river, Beroun gauging station (BB), (**b**) Labe river, Děčín gauging station (LD), (**c**) Morava river, Olomouc-Nové Sady gauging

station (MO), (**d**) Bečva river, Teplice nad Bečvou gauging station (BT), (**e**) Morava river, Kroměříž gauging station (MK), (**f**) Bystrzyca river, Krasków gauging station (BK), (**g**) Opawa river, Branice gauging station (OB), (**h**) Prudnik river, Prudnik gauging station (PP), (**i**) Ślęza river, Białobrzezie gauging station (SB), (**j**) Widawa river, Zbytowa gauging station (WZ).

The shape of the histograms of the empirical data for BB, LD, MO, BT, MK, BK, OB, PP, and SB of the data series is right-skewed. Only the histogram of the empirical data for WZ has a uniform shape (Figure 3). Histograms of empirical data (with the exception of the WZ data series) determine the best-fit (skewed-right) distributions from the Weibull, Gamma, and Log-Normal families. Hence, the 3LN, 3W, GEV, and 3PIII distributions were selected. In order to estimate the three parameters of these distributions, the Maximum Likelihood estimation was used. The estimated distributions are shown in Figure 3. For the WZ data series, we also decided to use the above-mentioned distributions.



Figure 3. Histograms showing the empirical, 3LN, GEV, 3W, and 3PIII distributions of the analyzed data series.

The plot in Figure 4 shows the CDF estimates of all the three-parameter distributions and the estimates of the empirical distribution function (ECDF). The CDF plot indicates that for the BB, the fit of 3PIII is poor compared to the ECDF estimate. The 3W distribution is also a poor fit. In turn, the other distributions seem to be quite close to the ECDF estimate. This shows that there can be more than one distribution well fitted to the empirical data. The plots for the LD, MK, SB, and WZ data series reveal that all the distributions (LN3, 3W, GEV, and 3PIII) fit the ECDF almost identically (Figure 4). The use of 3PIII for non-stationary flood peak time series leads to an underestimation of variability. This has been shown for both the MO and the BK data series. In this case, the other distributions (LN3, GEV, 3W) seem to be quite close to the ECDF estimate (Figure 4). The plot for two samples, BT and OB, shows that LN3 and GEV are closer to the ECDF estimate than 3PIII and 3W (Figure 4). In turn, for the maximum flow of the PP data series, GEV, 3W, and 3PIII are a poor fit as compared to the ECDF estimate. Hence, the best fit is obtained for the 3LN distribution (Figure 4).

Comparing the shift of the ECDF to the proposed distributions in the upper and lower tail, one can see that a single theoretical distribution will not always provide a good fit in both parts of the tail (Figure 4). An analysis of two consecutive samples MO and BT reveals that while all the four distributions (GEV, LN3, W3, 3PIII) seem to be appropriate in terms of fit to ECDF from the 95th to the 99th percentile (upper tail), only the 3W distribution provides the best fit on the lower tail section (Figure 4).

Another sample, the MK data series, shows that all the distributions considered seem to be quite close to the ECDF from the 95th to the 99th percentile (upper tail), but only the 3PIII fits the ECDF on its lower tail (Figure 4). In turn, for BB and LD, the GEV distribution seems to be quite close to the ECDF estimate from the 95th to the 99th percentile (upper tail). Next, in the lower tail for BB and LD (only from the 1st to the 10th percentile), the GEV seems to be quite close to the ECDF estimate (Figure 4). The long time series, as in the case of the BB, LD, MO, BT, MK, BK, and PP samples, have a smaller spread (Figure 4). In turn, the OB, SB, and WZ data series have a greater spread. In the upper tail of the ECDF, the 3W distribution for the BK and WZ seems to provide the best fit to the empirical data. In the case of OB, all of the distributions considered seem to fit the ECDF from the 98th to the 99th percentile. In turn, from the 95th to the 90th percentile, the GEV seems to be the best possible choice in terms of fit to ECDF (Figure 4). In the case of two samples, PP and SB, different distributions are fitted to the ECDF, depending on the percentile value. For the PP data series, the LN3 seems to fit the ECDF from the 90th to the 98th percentiles. In turn, for the 99th percentile, the 3W seems to be the best possible choice in terms of fit to ECDF (Figure 4). While the 3PIII distribution seems to fit the ECDF for the SB sample from the 90th to the 99th percentile, on the 99th percentile, the 3W seems to be the best choice in terms of fit to ECDF (Figure 4). In the lower tail of the ECDF, the 3W, GEV, LN3, 3PIII, and 3W distributions seem to provide the best fit to the empirical data for the BK, OB, PP, SB, and WZ, respectively (Figure 4).

Despite some ambiguities in indicating the best candidates, the CDF plots allowed for an analysis of the lower and upper tail of the distributions.

The empirical Probability Density Function superimposed on the PDF of the four theoretical distributions is shown in Figure 5. For all the samples, the empirical density is positively skewed. For the BB data series, the 3W and GEV distributions seem to be most appropriate for the empirical data series (Figure 5). In turn, for LD, SB, and WZ, all of the distributions (GEV, LN3, 3PIII, and 3W) seem to be the best possible choices for empirical data series (Figure 5). For the BK data series, only the GEV gives the best fit, and for MO and MK, the 3W distribution seems to be most appropriate for empirical data series (Figure 5). For the BT data series, the 3PIII and 3W differ significantly from the LN3 and GEV in the central portion of the range of Y values. Hence, the latter two distributions fit the data series almost identically. For the OB and PP samples, the LN3 and GEV seem to be the best possible choices appropriate for the empirical data series (Figure 5).



Figure 4. The empirical, 3LN, GEV, 3W, and 3PIII Cumulative Distribution Functions of the analyzed data series.

The *p*-value of the Anderson–Darling test shows that the 3PIII distribution for BB, BK, BT, MO, OB, and PP, the 3LN distribution for MO and BK, and the GEV distribution for WZ should be rejected (p < 5%). Rejected distributions were not included in the GOF test in Appendix B (Table A2).



Figure 5. The Empirical Probability Density Function and the Probability Density Function of 3LN, GEV, 3W, and 3PIII.

The CDF plot indicated that for a large number of samples, the GEV and the 3LN seem to be quite close to the ECDF (Figure 4). In turn, an analysis of the PDF plot (Figure 5) shows that 3W seems to be the best possible choice suited for most empirical data series (Figure 5). However, the performance rank score indicated that GEV had the most points (Table 3). Nevertheless, the 3LN distribution had a high score in two accuracy measure methods: MAE and RMSE. In turn, GEV obtained the highest number of points in the rank scores, which took into account the Anderson–Darling test, MARE, and information criteria (Table 3). While the 3W and GEV distributions prepared for the MO and BT data series,

respectively obtained the most points in all the analyzed methods and criteria (Table 3), for the MK data series, the Anderson–Darling test indicated that the 3W was the best fit, which was consistent with the PDF plot (Figure 5, Table 3). In turn, the RMSE pointed at the GEV distribution as the best fit. Despite this, it was the 3PIII that obtained the most points in the remaining accuracy measures methods, as well as in the AIC and BIC criteria.

Data Series	Distribution	AD Amax	MARE	MAE	RMSE	AIC	BIC	Total Rank
BB	GEV	3	3	3	3	3	3	18
	LN3	2	2	2	2	2	2	12
	3W	1	1	1	1	1	1	6
	GEV	2	4	3	3	4	4	20
ID	LN3	1	3	4	4	2	2	16
LD	3W	3	1	1	1	1	1	8
	PIII	4	2	2	2	3	3	16
MO	GEV	1	1	1	2	1	1	7
MO	3W	2	2	2	1	2	2	11
	GEV	3	3	3	3	3	3	18
BT	LN3	1	2	2	2	2	2	11
	3W	2	1	1	1	1	1	7
	GEV	1	1	1	4	2	2	11
MK	LN3	2	2	2	3	3	3	15
MIK	3W	4	3	3	1	1	1	13
	PIII	3	4	4	2	4	4	21
BV	GEV	2	2	1	1	1	1	8
BK	3W	1	1	2	2	2	2	10
	GEV	3	3	3	3	3	3	18
OB	LN3	1	2	2	1	2	2	10
	3W	2	1	1	2	1	1	8
	GEV	2	1	1	1	1	1	7
PP	LN3	3	3	3	3	2	2	16
	3W	1	2	2	2	3	3	13
	GEV	2	1	2	1	1	1	8
CP	LN3	3	3	1	2	2	2	13
30	3W	1	2	3	4	3	3	16
	PIII	4	4	3	3	3	3	20
	LN3	2	1	1	1	1	1	7
WZ	3W	1	3	3	3	3	3	16
	PIII	3	2	2	2	2	2	13

 Table 3. Distribution rank score in the Goodness-of-Fit tests.

On the one hand, GEV is the best fit for the BK data series according to the Anderson– Darling test and MARE, but on the other, 3W is suggested by the MAE, RMSE, AIC, and BIC. The latter distribution obtained a high score (Table 3).

The CDF and PDF plots indicated that GEV and 3LN are the best possible choices suited for the empirical OB data series (Figures 4 and 5). Nevertheless, the GEV distribution

prepared for this data series obtained the most points in all the analyzed methods and criteria (Table 3). The 3LN prepared for the PP data series was the best fit according to both the Anderson–Darling test and the accuracy measure methods (Table 3). It was consistent with the CDF plot (Figure 4). This distribution had a high score and is the best choice. Nevertheless, according to the AIC and BIC criterion, 3W is the best quality model out of all the other models. The 3W and 3PIII distributions prepared for the SB data series have a high score. However, in the overall rank, 3PIII is best. The 3W distribution prepared for the WZ data series has a high score. However, according to the Anderson–Darling test, the 3PIII distribution provides the best fit for the empirical data.

In this study, the sample size, catchment area, and flow regime do not seem to be important as factors that would favor a certain distribution.

Comparing the results obtained by the rank score of distributions (Table 3) and the plot showing the change points (Figure 2), we found that the 3PIII distribution provided the best fit only for two of the empirical data: MK and SB. For these two data series, no change points were detected, but a trend was found for MK. The remaining distributions (GEV, 3W, and 3PIII) fit the empirical non-stationary series. Moreover, for the LD and WZ data series, the Mann–Kendall test rejected the hypothesis of there being no trend. Despite this, the GEV and W distributions proved to be the best fit, respectively.

4. Discussion

In the present study, we analyzed the rivers on which the reservoirs exist or are planned. One river was chosen because of the flood protection in its valley. The analysis of the possible distributions can be widely used for flood protection and engineering calculations.

The study aimed to compare three-parameter distributions in controlled catchments for a stationary and non-stationary data series with the use of PDF plots, CDF plots, CDF plots (only lower tail of the distributions), CDF plots (only upper tail of the distributions), and a GOF test.

Selection of the optimal estimated parameters of the distribution requires prior verification of the data series. The following verification methods were used in the present study: change point analysis, temporal trend analysis, test for randomness. Similar methods were used in other research [4,20,70]. Some researchers used only one method [11]. Other researchers assumed that all the data series were not homogeneous [13].

In this study, the change point analysis indicated that only two in 10 samples have no change point. In another study, a small number of samples (18 out of 50) exhibited a significant abrupt change in the mean [53]. The Von Neumann ratio test, which was used in the study of the Litija gauging station (in Slovenia), showed that all the samples were randomized [20]. This is in line with [4] and with the present study. In the present study, the Mann–Kendall test showed that the LD and WZ samples had a negative trend and the Kendall test showed that, additionally, the MK sample had a negative trend, too. Only the SB was stationary. In other study, a negative trend was also recorded for the LD sample for the period 1851–2010 [70]. An analysis of the middle part of the River Morava—the MO and MK samples—revealed a negative trend [33]. The Widawa profile Zbytowa was not homogeneous [13]. Different approaches to the temporal trend have been recorded in many studies. In the samples of the Litija gauging station (in Slovenia), temporal trends were observed. In this case, all of the data series were used in the FFA. In turn, in the Rio Grande do Sul State (in Brazil), only seven out of 113 data series presented a significant monotone trend. They were not used for the investigation [11]. The study of the Upper Odra Basin assumed that all the data series were not homogeneous. Despite this, the scientists performed FFA using simple distributions. In the studies mentioned above, one or few methods were used. In our opinion, there is a need to verify the data series using simultaneously change point analysis, temporal trend analysis, test for randomness before the FFA, and non-stationary FFA analysis.

After verifying the data series, the next important step should be to choose the appropriate distribution. The determination of skewness and kurtosis makes it possible to determine the family of the distribution. The positive skewness and kurtosis of this sample (close to three) suggested that right-skewed distributions such as Normal, Log-Normal, Gamma, and Weibull should be considered for possible model distribution candidates [4]. In this study, the data series with a skewness of 0.30 means it is fairly symmetrical, a skewness of 0.92 means it is moderately skewed, and the skewness above 1.0 means it is highly skewed (Table 2). Moreover, a kurtosis greater than 3 has heavier tails than the normal distribution, while those less than 3 have lighter tails than the normal distribution (Table 2).

In our opinion, the use of histograms for empirical data easily allows for such an assessment (Figure 3). We found that histograms of the majority empirical data series determine the best-fit (skewed-right) distributions from the Weibull, Gamma, and Log-Normal families. In other studies, the histograms were used to illustrate fitted distributions to the empirical data series [4,5].

The sample size of the time series and their catchment area do not seem to be important as factors favoring a particular distribution in the study of the Torne River (Sweden) [4]. This is in line our own results. In turn, the elevation of the gauging station, the length of the straight line from the source to the gauging station, and the drainage area were the most important variables in explaining the peak flows [14].

In this study, the MLE method was used for stationary data and non-stationary data. The MLE is effective for longer time series [71] such as those analyzed in our study. The MLE has been successfully applied to the PIII distribution for a non-stationary flow data series [20]. In the study of the Praha gauging station on the Vltava river, the MLE was used to estimate 3LN [27]. The GEV parameters can be also estimated by means of the MLE [33]. In our opinion, there are not enough studies with MLE to estimate the three-parameter distributions for non-stationary flow data series. There are several other methods for estimating distribution parameters, but they have some limitations compared to MLE. For instance, the GAMLSS software cannot implement the GEV and LN3 distributions, and the TS method proved to be the best estimator for small sample sizes and trends in the mean values [30]. Based on our calculations, the MLE simplifies the procedure of determining theoretical distributions.

When comparing empirical and theoretical data, it is sometimes difficult to choose a single distribution. This was demonstrated by the analyses of the CDF and PDF plots (Figures 4 and 5). Then, the GOF test is used. In the present study, when analyzing the CDF plot and GOF tests, we noticed that the AD test did not always indicate the best distribution in comparison to the accuracy measures. Some researchers [11,13] used statistical tests and GOF as the accuracy measure. The accuracy measures and GOF test can be applied separately [4] In addition, the AIC can be added to the GOF test [5,20].

In the present study, the AD test was used because it assigns more weight to the distribution tails [11]. In another study, both the AD test and the K-S test selected the same distribution—3PIII—as the best fit for empirical data. The parameters of this distribution were estimated using the MLE [20]. Lastly, it was found that all analyzed distributions obtained identical results when comparing the tests: AD, K-S test, and Cramer-von Mises for probable maximum flows [5].

After estimating the distribution parameters using MLE, it is recommended to use the AIC identification [17]. The MLE-estimated parameters for the 3PIII and the GEV distributions at the Litija gauging station of the Sava River showed that the AIC values were 187.19 and 193.76, respectively [20]. These values were lower than those obtained in the studies of rivers in the Polish-Czech borderland. In turn, for the Tana river at the Garissa gauging station, the 3PIII and 3LN distributions have the lowest AIC and BIC. The values were 1083.1 (AIC), 1087.7 (BIC), 1083.7 (AIC), and 1088.4 (BIC), respectively. This was the reason for selecting these distributions as the best models. [5]. However, in the present study, the AIC and BIC tests were not always consistent with the results of the AD test and the accuracy measure method.

According to our research results and other studies using the GOF test in many cases, it may turn out to be insufficient. Different statistical tests, accuracy measures, and informa-

tion criteria give different solutions, not always indicating one distribution [5,11,13,20,25]. Therefore, a method should be used that allows the elimination of mismatched distributions before the GOF test is performed. Such a solution is to analyze the CDF plot. The quantile estimation of the selected upper tail of the distribution is very often performed due to the use of extreme flow values in the design and flood risk assessment. Such analyses were performed by [4,11,19,20,28,30]. The CDF analysis was presented by selecting the distributions closest to the empirical data [5]. In addition, in the Rhine river study, the ECDF plot was analyzed for the flood wavelength above the 95th quantile threshold of individual time series discharge using the storm surge model and two hydrological river discharge models [2]. It is worth noting that the lower tail quantiles are also used in the engineering practice to calculate the occurrence of floods in agricultural areas, in rainwater drainage, and drainage construction. In the present study, we show that the analysis of the CDF and ECDF graphs allows one to indicate the theoretical distributions closest to the empirical distribution in both the upper and lower tail (Table 4). If we need quantiles from only one tail (either the lower or the upper one), we can choose the theoretical distributions most suited for the empirical data from this tail before performing the Goodness-of-Fit test.

Table 4. Best-fit distributions using graphical (CDF plot) and analytical (GOF test) methods.

Data Series	CDF (Distribution)	CDF (Only Lower Tail)	CDF (Only Upper Tail)	GOF
BB	GEV, LN3	GEV	GEV	GEV
LD	GEV, LN3, PIII, 3W	GEV	GEV	GEV
МО	GEV, 3W	3W	GEV, LN3, PIII, 3W	3W
BT	GEV, LN3	3W	GEV, LN3, PIII, 3W	GEV
МК	GEV, LN3, PIII, 3W	PIII	GEV, LN3, PIII, 3W	PIII
ВК	GEV, 3W	3W	3W	3W
OB	GEV, LN3	GEV	GEV, LN3, PIII, 3W	GEV
PP	LN3	LN3	LN3, 3W	LN3
SB	GEV, LN3, PIII, 3W	PIII	PIII	PIII
WZ	LN3, PIII, 3W	3W	3W	3W

Based on our research and analyses of other scientific works, a procedure has been proposed (Figure 6). The adopted methodology, not limited only to the FFA, is a proposal that can be used in the scientific and practical flood analysis as well as in engineering design. In water management, this should be helpful in the analysis of various flow values with a given probability of exceeding.

The novelty of this work is as follows:

- Proposing to standardize the verification of data series before performing the FFA and non-stationary FFA procedures;
- Histograms of the empirical data series determine the family of the distributions; the use of histograms easily allows for such an assessment;
- The MLE method allows one to simplify the analysis whenever there is a series of data, both stationary and non-stationary;
- Combining the graphical method (CDF plot) and the analytical method (GOF test);
- Our results will find the optimal three-parameter distribution for a given data series in the upper and/or the lower tail.



Figure 6. Workflow for determining theoretical distributions for a stationary and non-stationary data series.

5. Conclusions

The study aimed to compare three-parameter distributions in controlled catchments for a stationary and non-stationary data series with the use of PDF plots, CDF plots, CDF plots (for both lower and upper tail of the distributions), and a GOF test.

The data was analyzed according to the following procedure. (1) The data series were verified. (2) The stationarity of the data series was determined. (3) Based on skewness and kurtosis analyses as well as histograms, families of distributions were selected. (4) The parameters of distributions with the use of MLE were estimated. (5) The PDF plots and CDF plots were analyzed. (6) The fit of theoretical and empirical data in the lower and upper tail of the distributions was analyzed. (7) The GOF test was performed, and a theoretical distribution was chosen for each analyzed sample.

We found that our new procedure could be broadly applicable in the determination of the three-parameter distributions in controlled catchments for a stationary and nonstationary data series. The methodology used in this work, i.e., making use of the MLE, allows one to simplify the analyses when there is a series of data, both stationary and non-stationary.

The analyses performed in our article will be used in flood study, water resource planning, and the designing of hydraulic structures.

The most significant novelty in our research is the standardization and development of a new procedure for a stationary and non-stationary data series, taking into account to read a specific value of the maximum flow with a given exceedance probability from the lower or upper tail. From the point of view of engineering practice, this is extremely important. It determines the optimal choice of the theoretical distribution that can be used, for example, in the design of weirs in rural areas (lower quantiles) or in the design of hydrotechnical structures in areas at risk of flooding (upper quantiles).

Further research will evaluate the influence of natural and anthropogenic causes and the influence of seasonal changes on the parameters of the empirical distribution.

Author Contributions: Conceptualization, Ł.G., M.W., J.P.S. and J.P.; methodology, Ł.G.; software, Ł.G.; validation, M.W., J.P.S. and J.P.; formal analysis, Ł.G.; investigation, Ł.G. and M.W.; resources, Ł.G., J.P.S. and J.P.; data curation, Ł.G.; writing—original draft preparation, Ł.G., M.W. and P.T.; writing—review and editing, Ł.G. and P.T.; visualization, Ł.G. and P.T.; supervision, M.W., J.P.S. and J.P.; project administration, Ł.G.; funding acquisition, Ł.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by Wrocław University of Environmental and Life Sciences funding for an Innovative researcher (N060/0006/20).

Data Availability Statement: The dataset analyzed in this study is publicly available. These data can be found at the Czech Hydrometeorological Institute (https://www.chmi.cz/, accessed on 3 December 2020) and at the Institute of Meteorology and Water Management—National Research Institute (IMWM-NRI) (https://danepubliczne.imgw.pl, accessed on 3 December 2020). The data presented in this study are also available on request from the corresponding author.

Acknowledgments: The authors would like to express their sincere gratitude to the Czech Hydrometeorological Institute and the Institute of Meteorology and Water Management—National Research Institute for the release of the flow data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Figure A1 and Table A1 show the variability of long-term monthly flow rates k and selected characteristics of the hydrological regime in the studied rivers in Poland and the Czech Republic.



Figure A1. Cont.



Figure A1. Variability of long-term monthly flow rates k in the studied rivers in Poland and the Czech Republic.

Table A1. Selected characteristics of the hydrological regime in the studied rivers of Poland and the
Czech Republic.

River, Station	Kr	Cm	Qmean _{XII-V} and Qmean _y (Qmean _a)	Regime-Parde	Regime– Dynowska
Berounka, Beroun	3.80 (average balanced flows)	36.53% (slightly unbalanced outflow)			
Labe, Decin	3.41 (average balanced flows)	33.68% (slightly unbalanced outflow)	_	simple	
Morava, Olomouc	4.10 (average balanced flows)	42.76% (slightly unbalanced outflow)	 Qmean_{XII-V} > Qmean_y (lowland and highland streams) 	snow–rain, mountain variety	nival
Becva, Teplice	3.47 (average balanced flows)	39.40% (slightly unbalanced outflow)	_		
Morava, Kromeriz	3.78 (average balanced flows)	40.16% (slightly unbalanced outflow)	_		
Bystrzyca, Kraskow	2.91 (very balanced alpine flows)	30.5% (slightly unbalanced outflow)	Qmean _{XII-V} > Qmean _y		nival–
Opawa, Branice	3.51 (average balanced flows)	34.34% (slightly unbalanced outflow)	(lowland and highland streams)	composite	pluvial
Prudnik, Prudnik	2.79 (very balanced alpine flows)	25.38% (balanced outflow)	Qmean _{XII-V} < Qmean _y (alpine streams)	regime	pluvial– nival
Sleza, Bialobrzezie	1.68 (very balanced alpine flows)	19.75% (balanced outflow)	Qmean _{XII-V} > Qmean _y		nival– pluvial
Widawa, Zbytowa	5.03 (average balanced flows)	47.5% (slightly unbalanced outflow)	(lowland and highland streams)	composite rain, oceanic variety	nival

Appendix B

Table A2 showing Goodness-of-Fit of the Anderson–Darling (AD) test and accuracy measures, and AIC and BIC criterion of the estimated distributions of the analyzed data series.

Data Series	Distribution	AD Amax	AD <i>p</i> .Value	MARE	MAE	RMSE	AIC	BIC
	GEV	1.05	0.972	3.81	15.09	61.74	995.4	1002.5
BB	LN3	1.53	0.812	4.15	18.20	75.80	995.7	1002.7
	Weibull	2.36	0.429	9.25	31.82	89.38	1002.9	1009.9
	GEV	3.10	0.262	3.93	49.98	79.88	2107.5	2116.1
	LN3	3.93	0.112	4.62	45.20	61.23	2113.9	2122.6
LD	Weibull	2.18	0.604	7.22	92.04	141.40	2116.2	2124.9
	PIII	1.35	0.946	4.77	65.44	117.04	2110.1	2118.7
	GEV	2.98	0.255	3.89	7.75	17.21	1129.6	1137.3
MO	Weibull	2.48	0.413	3.40	7.74	24.60	1129.3	1137.1
	GEV	1.69	0.773	4.92	11.08	29.58	1088.2	1095.8
BT	LN3	2.98	0.256	5.75	12.95	34.28	1090.8	1098.4
	Weibull	1.86	0.694	11.21	21.12	40.90	1104.8	1112.4
МК	GEV	2.74	0.323	2.62	9.46	21.10	1300.2	1308.1
	LN3	2.33	0.473	2.50	9.13	21.35	1299.5	1307.4
	Weibull	1.92	0.664	2.26	8.66	26.85	1301.2	1309.1
	PIII	2.00	0.622	2.06	7.77	23.67	1299.2	1307.1
DI	GEV	1.83	0.626	9.98	12.78	40.38	709.2	715.9
DK	Weibull	2.25	0.435	13.20	8.36	15.69	706.5	713.2
	GEV	2.52	0.302	9.54	9.48	30.28	488.4	494.0
OB	LN3	1.94	0.524	10.46	10.46	31.90	488.9	494.6
	Weibull	2.21	0.409	17.07	13.50	31.58	492.7	498.3
	GEV	2.15	0.477	12.98	6.64	22.17	599.9	606.4
PP	LN3	2.14	0.483	10.59	3.59	6.33	597.1	603.6
	Weibull	3.16	0.175	11.05	3.61	7.01	593.9	600.4
	GEV	3.33	0.146	13.44	1.32	3.10	416.1	422.8
CD	LN3	1.92	0.580	12.20	1.12	2.50	413.8	420.5
50	Weibull	3.77	0.092	12.39	0.87	1.25	411.1	417.8
	PIII	1.13	0.939	12.01	0.87	1.33	411.1	417.8
	LN3	1.53	0.729	9.63	1.87	2.87	366.9	372.6
WZ	Weibull	2.09	0.458	7.38	1.18	1.44	361.0	366.6
	PIII	1.50	0.744	7.68	1.37	1.86	363.5	369.2

Table A2. Goodness-of-Fit of the Anderson–Darling (AD) test and accuracy measures, and AIC and BIC criterion of the estimated distributions of the analyzed data series.

References

- 1. Seckin, N.; Yurtal, R.; Haktanir, T.; Dogan, A. Comparison of probability weighted moments and maximum likelihood methods used in flood frequency analysis for Ceyhan river basin. *Arab. J. Sci. Eng.* **2010**, *35*, 49.
- 2. Khanal, S.; Ridder, N.; de Vries, H.; Terink, W.; van den Hurk, B. Storm Surge and Extreme River Discharge: A Compound Event Analysis Using Ensemble Impact Modeling. *Front. Earth Sci.* **2019**, *7*, 224. [CrossRef]
- 3. Buchlák, J.; Matějka, J.; Ryjáček, P.; Bílý, P.; Procházka, J.; Pollert, J.; Fabel, J. Experimental verification of functionality of fibre-reinforced concrete submersible piers. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *596*, 012029. [CrossRef]
- 4. Hassan, M.U.; Hayat, O.; Noreen, Z. Selecting the best probability distribution for at-site flood frequency analysis, a study of Torne River. *SN Appl. Sci.* **2019**, *1*, 1629. [CrossRef]
- 5. Langat, P.K.; Kumar, L.; Koech, R. Identification of the Most Suitable Probability Distribution Models for Maximum, Minimum, and Mean Streamflow. *Water* **2019**, *11*, 734. [CrossRef]

- 6. Myronidis, D.; Stathis, D.; Sapountzis, M. Post-Evaluation of Flood Hazards Induced by Former Artificial Interventions along a Coastal Mediterranean Settlement. *J. Hydrol. Eng.* **2016**, *21*, 05016022. [CrossRef]
- Sweet, W.V.; Genz, A.S.; Obeysekera, J.; Marra, J.J. A Regional Frequency Analysis of Tide Gauges to Assess Pacific Coast Flood Risk. Front. Mar. Sci. 2020, 7, 581769. [CrossRef]
- Dunne, T.; Leopold, L.B. (Eds.) Calculation of flood hazard. In *Water in Environmental Planning*; W.H. Freeman: San Francisco, CA, USA, 1978; pp. 279–391.
- 9. Alila, Y.; Mtiraoui, A. Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques. *Hydrol. Process.* 2002, *16*, 1065–1084. [CrossRef]
- 10. Katz, R.W.; Parlange, M.B.; Naveau, P. Statistics of extremes in hydrology. Adv. Water Resour. 2002, 25, 1287–1304. [CrossRef]
- Cassalho, F.; Beskow, S.; de Mello, C.R.; de Moura, M.M.; Kerstner, L.; Ávila, L.F. At-Site Flood Frequency Analysis Coupled with Multi-parameter Probability Distributions. *Water Resour. Manag.* 2018, *32*, 285–300. [CrossRef]
- Młyński, D.; Petroselli, A.; Wałęga, A. Flood frequency analysis by an event-based rainfall–runoff model in selected catchments of southern Poland. *Soil Water Res.* 2018, 13, 170–176. [CrossRef]
- 13. Szulczewski, W.; Jakubowski, W. The Application of Mixture Distribution for the Estimation of Extreme Floods in Controlled Catchment Basins. *Water Resour. Manag.* **2018**, *32*, 3519–3534. [CrossRef]
- 14. Myronidis, D.; Ivanova, E. Generating Regional Models for Estimating the Peak Flows and Environmental Flows Magnitude for the Bulgarian-Greek Rhodope Mountain Range Torrential Watersheds. *Water* **2020**, *12*, 784. [CrossRef]
- 15. Rahman, A.; Zaman, M.A.; Haddad, K.; Adlouni, S.E.; Zhang, C. Applicability of Wakeby distribution in flood frequency analysis: A case study for eastern Australia. *Hydrol. Process.* **2015**, *29*, 602–614. [CrossRef]
- Xiong, L.; Du, T.; Xu, C.Y.; Guo, S.; Jiang, C.; Gippel, C.J. Non-stationary annual maximum flood frequency analysis using the norming constants method to consider non-stationarity in the annual daily flow series. *Water Resour. Manag.* 2015, 29, 3615–3633. [CrossRef]
- 17. Strupczewski, W.; Singh, V.; Feluch, W. Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. *J. Hydrol.* **2001**, *248*, 123–142. [CrossRef]
- Strupczewski, W.G.; Kochanek, K.; Bogdanowicz, E.; Markiewicz, I.; Feluch, W. Comparison of Two Nonstationary Flood Frequency Analysis Methods within the Context of the Variable Regime in the Representative Polish Rivers. *Acta Geophys.* 2016, 64, 206–236. [CrossRef]
- 19. Markiewicz, I.; Strupczewski, W.G.; Kochanek, K. On accuracy of upper quantiles estimation. *Hydrol. Earth Syst. Sci.* 2010, 14, 2167–2175. [CrossRef]
- 20. Bezak, N.; Brilly, M.; Šraj, M. Comparison between the peaks over threshold method and the annual maximum method for flood frequency analyses. *Hydrol. Sci. J.* **2014**, *59*, 959–977. [CrossRef]
- Kidson, R.; Richards, K.S. Flood frequency analysis: Assumptions and alternatives. Prog. Phys. Geogr. 2005, 29, 392–410. [CrossRef]
- Madsen, H.; Rasmussen, P.F.; Rosbjerg, D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modelling. *Water Resour. Res.* 1997, 33, 747–757. [CrossRef]
- 23. Svensson, C.; Kundzewicz, Z.W.; Maurer, T. Trend detection in river flow 697 series: 2. Flood and low-flow index series. *J. Hydrol. Sci.* 2005, *50*, 811–824. [CrossRef]
- 24. Gharib, A.; Davies, E.G.R.; Goss, G.G.; Faramarzi, M. Assessment of the Combined Effects of Threshold Selection and Parameter Estimation of Generalized Pareto Distribution with Applications to Flood Frequency Analysis. *Water* **2017**, *9*, 692. [CrossRef]
- 25. Gruss, Ł.; Wiatkowski, M.; Buta, B.; Tomczyk, P. Verification of the Methods for Calculating the Probable Maximum Flow in the Widawa River in the Aspect of Water Management in the Michalice Reservoir. *Annu. Set Environ. Prot.* **2019**, *21*, 566–585.
- Holický, M.; Jung, K.; Sýkora, M. Assessment of Extreme Discharges of the Vltava River in Prague. In *Flood Recovery, Innovation and Response*; Proverbs, D., Brebbia, C.A., Penning Rowsell, E., Eds.; WIT Press: Southampton, UK, 2008; pp. 105–112. [CrossRef]
- Holický, M.; Sykora, M. Assessment of flooding risk to cultural heritage in historic sites. J. Perform. Constr. Facil. 2010, 24, 432–438. [CrossRef]
- Markiewicz, I.; Strupczewski, W.G.; Bogdanowicz, E.; Kochanek, K. Generalized Exponential Distribution in Flood Frequency Analysis for Polish Rivers. *PLoS ONE* 2015, 10, e0143965. [CrossRef]
- Strupczewski, W.G.; Markiewicz, I. Initial study of two shape parameter flood frequency distributions. *Publs. Inst. Geophys. Pol.* Acad. Sc. 2006, 390, 147–154.
- 30. Debele, S.E.; Strupczewski, W.B.; Bogdanowicz, E.A. comparison of three approaches to non-stationary flood frequency analysis. *Acta Geophys.* **2017**, *65*, 863–883. [CrossRef]
- Tegegne, G.; Melesse, A.M.; Asfaw, D.H.; Worqlul, A.W. Flood Frequency Analyses over Different Basin Scales in the Blue Nile River Basin, Ethiopia. *Hydrology* 2020, 7, 44. [CrossRef]
- 32. Laio, F. Cramer–von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resour. Res.* **2004**, *40*, *9*. [CrossRef]
- Brázdil, R.; Chromá, K.; Řezníčková, L.; Valášek, H.; Dolák, L.; Stachoň, Z.; Soukalová, E.; Dobrovolný, P. The use of taxation records in assessing historical floods in South Moravia, Czech Republic. *Hydrol. Earth Syst. Sci.* 2014, 18, 3873–3889. [CrossRef]

- 34. Madsen, H.; Lawrence, D.; Lang, M.; Martinkova, M.; Kjeldsen, T.R. A Review of Applied Methods in Europe for Flood-Frequency Analysis in a Changing Environment. NERC/Centre for Ecology & Hydrology; European Cooperation in Science and Technology (COST): Brussels, Belgium, 2013.
- 35. Yiou, P.; Ribereau, P.; Naveau, P.; Nogaj, M.; Brázdil, R. Statistical analysis of floods in Bohemia (Czech Republic) since 1825. *Hydrol. Sci. J.* **2006**, *51*, 930–945. [CrossRef]
- 36. Wojarnik, K.; Iwaniak, E.; Wiatkowski, M.; Czamara, W. Influence of Restored External Spoil Tip of a Lignite Mine on the Discharge in a Cross-Border Watercourse (PL–CZ). *Annu. Set Environ. Prot.* **2019**, *21*, 343–363.
- Rutkowska, A.; Żelazny, M.; Kohnová, S.; Łyp, M.; Banasik, K. Regional L-Moment-Based Flood Frequency Analysis in the Upper Vistula River Basin, Poland. Pure Appl. Geophys. 2017, 174, 701–721. [CrossRef]
- 38. Młyński, D.; Wałęga, A.; Stachura, T.; Kaczor, G.A. New Empirical Approach to Calculating Flood Frequency in Ungauged Catchments: A Case Study of the Upper Vistula Basin, Poland. *Water* **2019**, *11*, 601. [CrossRef]
- 39. CHMI. Czech Hydrometeorological Institute. 2020. Available online: https://www.chmi.cz/ (accessed on 22 April 2020).
- 40. IMWM. Institute of Meteorology and Water Management—National Research Institute. 2020. Available online: https://danepubliczne.imgw.pl/ (accessed on 15 November 2020).
- 41. MAECR (The Ministry of Agriculture and the Environment). *General Areas Protected for Surface Water Accumulation and the Basic Principles of Its Utilization;* MAECR Press: Prague, Czech Republic, 2011.
- Brázdil, R.; Řezníčková, L.; Valášek, H.; Havlíček, M.; Dobrovolný, P.; Soukalová, E.; Řehánek, T.; Skokanová, H. Fluctuations of floods of the river Morava (Czech Republic) in the 1691–2009 period: Interactions of natural and anthropogenic factors. *Hydrol. Sci. J.* 2011, *56*, 468–485. [CrossRef]
- DZMiUW Wrocław. Small Water Retention Program in the Lower Silesian Voivodship. Study Prepared by Agricultural University of Wroclaw—Hydrological Process Modeling Center; Agricultural University of Wroclaw—Hydrological Process Modeling Center: Wrocław, Poland, 2006.
- 44. WZMiUW (Provincial Board of Land Reclamation and Water Facilities in Opole). *Program for the Construction of Small Retention Reservoirs in the Opolskie Voivodeship. Study Prepared by EMPEKO;* WZMiUW Press: Opole, Poland, 2002.
- Wiatkowski, M.; Wiatkowska, B.; Gruss, Ł.; Rosik-Dulewska, C.; Tomczyk, P.; Chłopek, C. Assessment of the possibility of implementing small retention reservoirs in terms of the need to increase water resources. *Arch. Environ. Prot.* 2021, 47, 80–100. [CrossRef]
- 46. Tomczyk, P.; Wiatkowski, M. The Effects of Hydropower Plants on the Physicochemical Parameters of the Bystrzyca River in Poland. *Energies* 2021, 14, 2075. [CrossRef]
- 47. Déry, S.J.; Stahl, K.; Moore, R.D.; Whitfield, P.H.; Menounos, B.; Burford, J.E. Detection of runoff timing changes in pluvial, nival, and glacial rivers of western Canada. *Water Resour. Res.* **2009**, 45, W04426. [CrossRef]
- Zeiringer, B.; Seliger, C.; Greimel, F.; Schmutz, S. River Hydrology, Flow Alteration and Environmental Flow. In *Riverine Ecosystem Management*; Schmutz, S., Sendzimir, J., Eds.; Springer Nature Switzerland AG: Cham, Switzerland, 2018; pp. 67–89. [CrossRef]
- 49. Aksamit, N.O.; Whitfield, P.H. Examining the pluvial to nival river regime spectrum using nonlinear methods: Minimum delay embedding dimension. *J. Hydrol.* **2019**, *572*, 851–868. [CrossRef]
- Kuriqi, A.; Pinheiro, A.N.; Sordo-Ward, A.; Garrote, L. Flow regime aspects in determining environmental flows and maximising energy production at run-of-river hydropower plants. *Appl. Energy* 2019, 256, 113980. [CrossRef]
- Poschlod, B.; Willkofer, F.; Ludwig, R. Impact of Climate Change on the Hydrological Regimes in Bavaria. Water 2020, 12, 1599. [CrossRef]
- 52. Wrzesiński, D.; Sobkowiak, L. Transformation of the Flow Regime of a Large Allochthonous River in Central Europe—An Example of the Vistula River in Poland. *Water* 2020, *12*, 507. [CrossRef]
- 53. Villarini, G.; Serinaldi, F.; Smith, J.A.; Krajewski, W.F. On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resour. Res.* 2009, 45, W08417. [CrossRef]
- 54. Dierauer, J.; Whitfield, P. *Daily Streamflow Trend and Change Point Screening 'FlowScreen' (Version 1.2.6)*; 2019; Available online: https://CRAN.R-project.org/package=FlowScreen (accessed on 2 December 2020).
- 55. R Core Team. R: A Language and Environment for Statistical Computing [Computer Software Manual, Version 4.0.5]. Vienna, Austria. Available online: http://www.R-project.org/ (accessed on 15 November 2020).
- 56. Mann, H.B. Non-parametric tests against trend. *Econometrica* 1945, 13, 245–259. [CrossRef]
- 57. Kendall, M.G.; Gibbons, J.D. Rank Correlation Methods, 5th ed.; Oxford University Press: New York, NY, USA, 1990.
- Myronidis, D.; Fotakis, D.; Ioannou, K.; Sgouropoulou, K. Comparison of ten notable meteorological drought indices on tracking the effect of drought on streamflow. *Hydrol. Sci. J.* 2018, 63, 2005–2019. [CrossRef]
- 59. Bartels, R. The rank version of von Neumann's ratio test for randomness. J. Am. Stat. Assoc. 1982, 77, 40–46. [CrossRef]
- Signorell, A.; Aho, K.; Alfons, A.; Anderegg, N.; Aragon, T.; Arachchige, C.; Arppe, A.; Baddeley, A.; Barton, K.; Bolker, B.; et al. DescTools: Tools for Descriptive Statistics. Version 0.99.40; 2021. Available online: https://CRAN.R-project.org/package= DescTools (accessed on 17 March 2021).
- 61. Hensel, T.-G.; Barkemeyer, D. Statistical Methods for Life Data Analysis 'Weibulltools' (Version 2.0.0). 2021. Available online: https://cran.r-project.org/web/packages/weibulltools/weibulltools.pdf (accessed on 17 March 2021).
- 62. Stephenson, A. Functions for Extreme Value Distributions 'Evd' (Version 2.3–3). 2018. Available online: https://CRAN.R-project. org/package=evd (accessed on 17 March 2021).

- 63. Becker, M.; Klößner, S. Pearson Distribution System 'PearsonDS' (Version 1.1). 2017. Available online: https://CRAN.R-project. org/package=PearsonDS (accessed on 17 March 2021).
- 64. Millard, S.P.; Kowarik, A. Package for Environmental Statistics, Including US EPA Guidance 'EnvStats' (Version 2.4.0). 2020. Available online: https://CRAN.R-project.org/package=EnvStats (accessed on 17 March 2021).
- 65. Teimouri, M.; Gupta, A.K. On the three-parameter Weibull distribution shape parameter estimation. *J. Data Sci.* **2013**, *11*, 403–414. [CrossRef]
- 66. Abida, H.; Ellouze, M. Probability distribution of flood flows in Tunisia. Hydrol. Earth Syst. Sci. 2008, 12, 703–714. [CrossRef]
- 67. Braun, H. A simple method for testing goodness-of-fit in the presence of nuisance parameters. *J. R. Stat. Soc.* **1980**, *42*, 53–63. [CrossRef]
- 68. Faraway, J.; Marsaglia, G.; Marsaglia, J.; Baddeley, A. Classical Goodness-of-Fit Tests for Univariate Distributions 'Goftest' (Version 1.2–2). 2019. Available online: https://CRAN.R-project.org/package=goftest (accessed on 17 December 2020).
- 69. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. Akaike Information Criterion Statistics, 1st ed.; Springer: Dordrecht, The Netherlands, 1986.
- Hall, J.; Arheimer, B.; Borga, M.; Brázdil, R.; Claps, P.; Kiss, A.; Kjeldsen, T.R.; Kriaučiūnienė, J.; Kundzewicz, Z.W.; Lang, M.; et al. Understanding flood regime changes in Europe: A state-of-the-art assessment. *Hydrol. Earth Syst. Sci.* 2014, 18, 2735–2772. [CrossRef]
- El Adlouni, S.; Ouarda, T.B.M.J.; Zhang, X.; Roy, R.; Bobee, B. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resour. Res.* 2007, 43, W03410. [CrossRef]