

Article

# Statistical Study of Rainfall Control: The Dagum Distribution and Applicability to the Southwest of Spain

Fernando López-Rodríguez <sup>1</sup>, Justo García-Sanz-Calcedo <sup>1,\*</sup> , Francisco J. Moral-García <sup>2</sup> and Antonio J. García-Conde <sup>3</sup>

<sup>1</sup> Engineering Projects Department, Universidad de Extremadura, Avda. Elvas s/n, 06071 Badajoz, Spain; ferlopez@unex.es

<sup>2</sup> Graphical Expression Department, Universidad de Extremadura, Avda. Elvas s/n, 06071 Badajoz, Spain; fjmoral@unex.es

<sup>3</sup> Infrastructure Department, Diputación de Badajoz, 06002 Badajoz, Spain; ajgarcia.atm@dip-badajoz.es

\* Correspondence: jgsanz@unex.es; Tel.: +34-924-289-000

Received: 1 January 2019; Accepted: 22 February 2019; Published: 4 March 2019



**Abstract:** It is of vital importance in statistical distributions to fit rainfall data to determine the maximum amount of rainfall expected for a specific hydraulic work. Otherwise, the hydraulic capacity study could be erroneous, with the tragic consequences that this would entail. This study aims to present the Dagum distribution as a new statistical tool to calculate rainfall in front of frequent statistical distributions such as Gumbel, Log-Pearson Type III, Gen Extreme Value (GEV) and SQRT-ET max. The study was performed by collecting annual rainfall data from 52 meteorological stations in the province of Badajoz (Spain), using the statistical goodness-of-fit tests of Anderson–Darling and Kolmogorov–Smirnov to establish the degree of fitness of the Dagum distribution, applied to the maximum annual rainfall series. The results show that this distribution obtained a flow 21.92% greater than that with the traditional distributions. Therefore, in the Southwest of Spain, the Dagum distribution fits better to the observed rainfall data than other common statistical distributions, with respect to precision and calculus of hydraulics works and river flood plains.

**Keywords:** hydrology; statistical distributions; hydrology projects; rainfall control

## 1. Introduction

River flooding and hydrologic studies are carried out ensuring that the waters will reach a certain maximum level during statistic rainfalls with a certain return period. The possible level of flooding and the waterworks are designed by means of different classical statistical distributions applied to rainfall, using a series of maximum annual recorded rainfall data.

Some models of future scenarios suggest that climate change will involve a significant modification in the distribution of extreme rainfall intensity [1].

There have been several significant flooding events in the southwest of Spain in the last few decades. Particularly in Badajoz (Spain) in the year 1997, there were even casualties as a result of the river Rivillas breaking its banks. Therefore, undoubtedly another solution must be sought to provide a better fit to the historical rainfall data than those options currently available [2].

According to the scientific literature the statistical distributions most commonly used in Europe and Spain are: The Gumbel distribution, developed by the German mathematician Gumbel [3] and later applied to hydrology [4]; the Log-Pearson Type III distribution [5,6] put forward by several authors for use in hydrology [7,8]; extreme values distribution (GEV) [9]; and the SQRT-ET max. distribution [10,11], which best fits the characteristics of Spanish rainfall.

The method most commonly used in Spain for the calculation of the flood flow ( $\text{m}^3/\text{s}$ ) for a specific return period with the corresponding calculation changes is regulated by Spanish Highway Instruction Norm 5.2-IC [12]. For the calculation of the flow this regulation proposes the use of the Rational Method [13], which is based on the fact that the duration of the rainfall is greater than or equal to the concentration time in the basin and, therefore, it determines maximum flow rate. The rational method is used in Spain for small basins. For large ones, with concentration times of over 6 hours or more than  $1 \text{ km}^2$  (and up to  $3000 \text{ km}^2$ ), the modified method [8] was used, referred to by hydrologists such as Ferrer (1993) [14]. Currently, with the development of the computational capacity, more accurate hydro-meteorological models are widely used (e.g., software Hydrologic Modeling System, HEC-HMS). In other words, both models are in use depending on the basin size or predictable concentration times of more than 6 hours.

Panagoulia and Caroni (2011) [15] investigated the fit of generalized extreme value (GEV) distributions to maximum precipitation over the Mesochora mountainous catchment in central-western Greece. They observed that the stationary model was adequate for historical data on annual daily maxima for 1972–1992 and also for  $1 \times \text{CO}_2$  for the period 1961–2000. However, the  $2 \times \text{CO}_2$  series for 2061–2100 required a cubic time trend in location to obtain a satisfactory fit.

Lang et al. (1999) [16] reviewed tests and methods useful for modeling the process of over-threshold values, the choice of the threshold level, the verification of the independence of the values and the stationarity of the process. The independence criteria for the extraction of the peak-over-threshold events were recently modified by Onyutha (2017, 2019) [17,18] in terms of the independency ratio, threshold, and the inter-event time. Despite the desirable properties of the peak-over-threshold in the analyses of extreme events, this study focused on the annual maximum series.

It must be pointed out that these distributions are limited, as they must rely on rainfall for the desired return period. Additionally, there are scarce statistical distributions currently used in hydrology. They only were conceived at the beginning of the last century and scientific methods have, without doubt, moved on since then. For example, it seems that: (i) the distribution of Gumbel in high return periods underestimates precipitation values, and creates uncertainty about the safety of works designed with this criterion; (ii) the Log-Pearson Type III distribution does not underestimate values, but has three estimable parameters; (iii) the SQRT-ET max distribution is perhaps more precise and easier to apply, although it also has estimable parameters; (iv) the GEV distribution takes into account the extreme values maximum annual rainfall of one day but not the intermediate values. Other distributions, though not considered in this study, are Pareto [19] and Generalized Logistic [20].

However, the Dagum distribution [21] presents all the requisites to be used in hydrological studies. For example, it is a continuous distribution of extreme and non-negative values of the continuous variable. The Dagum distribution is used in different scientific fields, such as econometrics, economic theory (particularly for model ranges as diverse as wealth distribution), the banking sector (by using the distribution methods of aggregate losses) and the insurance sector (applied to analyze solvency) [22].

This distribution began to be used mostly in latitudes different from Europe for the calculation of rainfall, Mielke and Johnson [23] verified that a distribution similar to that of Dagum is possible in meteorology and hydrology. Later we found studies by Alam S. et al. (2014) [24], Alam and Rahat (2015) [25] in rivers of Dhaka (Bangladesh), Crabbe (2014) [26] in climate change systems in Charlottetown (Canada) and Mayooran and Laheetharan (2014) [27] in Colombo (Sri Lanka), possibly as a consequence of the existing problems of rainfall and floods in those countries.

On the other hand, in the East of Spain, Pérez-Sánchez and Senent-Aparicio (2018) [28], studied 29 meteorological stations between 1993 and 2013, for a given daily rainfall, simulating scenarios—for all those stations—with several distributions: Burr, Dagum, error, generalized extreme value, generalized logistic, generalized Pareto, Gumbel Max, inverse Gaussian, Johnson SB, Log-Logistic, Log-Pearson 3, Triangular, Weibull, and Wakeby. They observed that only the series of annual maximum dry spells offer a good adjustment for all the weather stations, showing that the Wakeby distribution presented

the best results. Finally, Domma and Condino (2017) [29] carried out a simulation study that shows the good performance of the maximum likelihood estimators for finite samples.

However, Mayoaran and Laheetharan (2014) [30] used the same form of Dagum as in this work and compared it to other 44 different distributions. The parameters of the selected probability distributions were used to generate random numbers for both actual and estimated maximum daily precipitation. Log-Pearson 3 and Burr (4P) were found to be the best-fit probability model for the annual period and the first inter monsoon study period, respectively [31].

The transmuted Dagum model provides the broader range of hazard behavior than the Dagum model [32]. The parameters of the new model are estimated by maximum likelihood using Newton Raphson approach and the information matrix and confidence intervals are also obtained. Other simulation results showed that both the corrected Akaike information criterion and Bayesian information criterion (BIC) always detected nonstationary, but the BIC selected the correct model more often except in very small samples [33]. Simulation studies indicated that the bias corrected and accelerated (BCa) method is best overall for the extreme percentiles that are often the focus of interest [34].

Despite all the above, it is not understandable that such a small number of distributions are used in professional practice in Europe and especially in Spain, considering that there are other efficient distributions in the field of hydrology.

To demonstrate the effectiveness of the Dagum distribution, the adjustment of the statistical distribution to the observed maximum annual rainfall values will be confirmed using the Anderson–Darling [35] and Kolmogorov–Smirnov goodness-of-fit tests and comparing them with the other distributions.

A statistical distribution must provide as good fit to the rainfall data as possible, since the better the fit, the more precise the value for the calculated rainfall. It can be used in the sizing of waterworks and flood plains.

This study intends to introduce the Dagum distribution as a new statistical tool to calculate rainfall because it fits better to the observed rainfall data by testing it with a dataset from Spain.

## 2. Materials and Methods

Firstly, a review of the statistical distributions used in hydrology studies will be performed, mainly in Europe and Spain, including the distribution of Dagum, with its fundamental characteristics.

The demonstration of the validity of this distribution in the field of civil engineering is analyzed to check whether the Dagum distribution provides a better fit, according to the goodness-of-fit tests, to the maximum annual rainfall distributions than the fit given by the commonly used distributions in the province of Badajoz, using real maximum annual rainfall data from the meteorological stations in that province.

### 2.1. Statistical Distribution Functions

The current method used in Spain to find out the flow rate that allows to dimension waterworks or calculate a flood plain of a river is the Rational Method [13], (except for large basins size). This method transforms the statistical rainfall associated to a certain return period (mm) to a flow rate (L/s):

$$Q = c \times I \times A \quad (1)$$

where  $c$  is a constant called runoff coefficient;  $I$  is the maximum intensity of precipitation and  $A$  is the area of the basin.

According to Instruction 5.2-IC, variable  $I$  can be obtained from the IDF curves (intensity, duration, frequency), with the following expression:

$$I_i = \frac{P_d}{24} \times \left( \frac{I_i}{I_d} \right)^{\frac{28^{0.1} - D^{0.1}}{0.4}} \quad (2)$$

As it can be verified, the correct estimation of  $P_d$  is very important, since it is the maximum daily precipitation obtained through the series of daily rainfall recorded in rainfall stations.

The rainfall associated with the return period is currently calculated using the statistical distributions commonly used in hydrology: Gumbel, Log-Pearson Type III, SQRT-ET max and GEV. Thus, starting from a historical record obtained from the rainfall stations the maximum value of rainfall associated with a certain return period is determined (frequency).

The use of one of these statistical distributions is nearly always found in hydrological studies [12].

### 2.1.1. The Gumbel Distribution

According to Gumbel [36], the density function is:

$$F(x) = e^{-e^{-\alpha(x-\mu)}} \quad (3)$$

where  $x$  is the value of the random variable,  $F(x)$  represents the probability that rainfall is less than or equal to  $x$ , and  $\alpha$  and  $\mu$  are parameters of the fit which depend on the mean and standard deviation of the variable  $y_i$ , which in turn depends on the sample size.

$$y_i = -\text{Ln Ln} \left( \frac{N+1}{i} \right) \alpha = \frac{S_N}{S_x} \mu = \bar{x} - \bar{y}_N \frac{S_x}{S_N} \quad (4)$$

Being  $\bar{y}$  and  $S_N$  the mean and standard deviation of the variable  $y_i$ , respectively,  $\bar{x}$  and  $S_x$  the mean and standard deviation respectively of the sample from the data of daily maximum values of annual rainfall:

$$\bar{x} = \sum_{i=1}^N \frac{x_i}{N} \quad S_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}} \quad \bar{y}_N = \frac{\sum_{i=1}^N y_i}{N} \quad S_N = \sqrt{\frac{\sum_{i=1}^N (y_i - \bar{y}_N)^2}{N}} \quad (5)$$

The return period,  $T(x)$ , is related to the distribution function,  $F(x)$ , by Equation (6):

$$T(x) = \frac{1}{1 - F(x)} \quad (6)$$

After entering sample values, the analytical expression is reached Equation (7), after clearing  $x$ :

$$x = P_{24}^T = P_d = \mu - \frac{1}{\alpha} \left[ \text{Ln} \left( -\text{Ln} \frac{T(x) - 1}{T(x)} \right) \right] \quad (7)$$

where the expected daily maximum precipitation  $P_d$  is obtained for a given return period  $T(x)$ .

### 2.1.2. The Log-Pearson Type III Distribution

The Log-Pearson density function is:

$$f(x) = \frac{\lambda^\beta (y - \varepsilon)^{\beta-1} e^{-\lambda(y-\varepsilon)}}{x \Gamma(\beta)} \quad (8)$$

being  $y = \log(x)$  and  $\Gamma(\beta)$  the function Gamma and  $e$  the Euler's number.

$\beta$ ,  $\lambda$  and  $\varepsilon$  are the parameters of form, scale and position respectively and are drawn from the equation (9–11):

$$\beta = \left(\frac{2}{C_s}\right)^2 \tag{9}$$

$$C_s = \frac{n}{(n-1) \times (n-2) \times (S_{\log(x)})^3} \times \sum (\log(x) - \overline{\log(x)})^3 \tag{10}$$

$$\lambda = \frac{S_y}{\sqrt{\beta}} \tag{11}$$

The density function of this distribution is not integral, so it is resolved by parametric methods.

$$P_{24}^T = 10^{\log(x) + K \times S_{\log(x)}} \tag{12}$$

### 2.1.3. The Distribution of SQRT-ET Max

The density function is as follows:

$$F(x) = \exp\left[-k \times (1 + \sqrt{\alpha \times x}) \times \exp(\sqrt{-\alpha \times x})\right] \tag{13}$$

where  $F(x)$  is the probability that the value will be less than  $x$ , and  $k$  and  $\alpha$  are parameters to be estimated that depend on the mean and typical deviation of the data series.

### 2.1.4. The Gen Extreme Value (GEV)

The density function [10] is:

$$f(x) = \begin{cases} \frac{1}{\sigma} e^{-(1+kz)^{-1/k}} (1+k \times z)^{-1-k \times z}, & k \neq 0, \\ \frac{1}{\alpha} e^{-e^{-z}}, & k = 0 \end{cases} \tag{14}$$

where  $z = x - \mu/\alpha$ , and  $k$ ,  $\mu$  and  $\alpha$  are the parameters of form, position and scale.

### 2.1.5. The DAGUM Distribution

The Dagum distribution has long been used in different fields such as economics, econometrics [37] and social sciences. However, there are just a few applications found in hydrology. The importance of using this distribution in hydrology is both its adoptability to extreme data and similar capability to traditional distributions.

In probability theory, statistics and econometrics, the Dagum distribution is a continuous distribution with a probability distribution defined on real positive numbers. The Dagum distribution arose from several variations of a new model in the size distribution of personal incomes and is associated above all with the study of incomes. This distribution can be used for three parameters (Type I) and for four parameter (Type II). The density function is defined by:

$$F(x) = \frac{\alpha k \left(\frac{x-\gamma}{\beta}\right)^{\alpha k-1}}{\beta \left(1 + \left(\frac{x-\gamma}{\beta}\right)^\alpha\right)^{k+1}} \tag{15}$$

and the distribution function

$$f(x) = \left(\left(1 + \frac{x-\gamma}{\beta}\right)^\alpha\right)^{-k} \tag{16}$$

where  $k$  is a continuous shape parameter ( $k > 0$ ),  $a$  is continuous shape parameter ( $a > 0$ ),  $\beta$  is a continuous scale parameter ( $\beta > 0$ ) and  $\gamma$  is a continuous location parameter ( $\gamma = 0$  yields the three-parameter Dagum distribution) ( $\gamma \leq x \leq \infty$ ). Figure 1 shows the density function of Dagum.

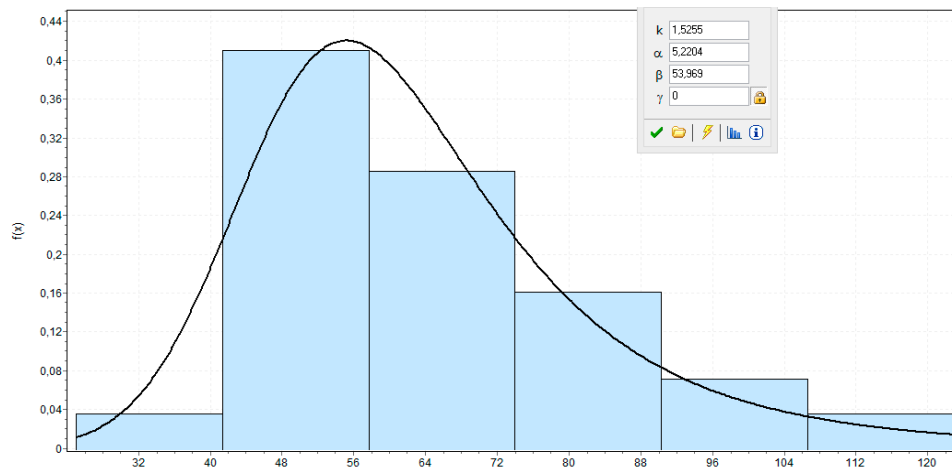


Figure 1. Density function of the Dagum distribution as a function of its parameters.

## 2.2. Tests of Goodness

For the adjustment of the distributions the EasyFit software [38] was used, which adjusts the probability laws to the rainy series and allows performing the goodness-of-fit tests by the Kolmogorov–Smirnov, Anderson–Darling and chi-square methods.

### 2.2.1. The Kolmogorov–Smirnov test

The Kolmogorov–Smirnov test [39] is a nonparametric, single sample, bi-sample and continuous test that proves particularly useful for large samples and is therefore optimal for the study [40].

The Kolmogorov–Smirnov test considers two hypotheses:

$$H_0: F(X) = F_s(X), H_1: F(X) \neq F_s(X) \tag{17}$$

where  $F(X)$  is the distribution function to be studied, and  $F_s(X)$  is the probability or theoretical proportion of values that must be less than or equal to  $x$  assuming the proposed hypothesis to be true.

Sample:  $n$  independent observations.

$$\text{Contrast statistic: } |D_n| = \text{maximum } |F_s(X) - S(X)| \tag{18}$$

### 2.2.2. The Anderson–Darling Test

The Anderson–Darling test has been widely used in hydrology due to its reliability in comparison with other tests and its common use in samples with pronounced tails. This test is very interesting compared to commonly used tests when faced with a variety of hydraulic engineering alternatives [41].

The Anderson–Darling test [42] uses the following formulation as a test statistic:

$$A_2 = -N - S \tag{19}$$

$$S = \sum_{i=1}^N \frac{(2i-1)}{N} \ln(F(x)_i + \ln(1 - F(x)_{N+1-i})) \tag{20}$$

where  $A_2$  is the test statistic,  $N$  is the sample size, and  $F(x)$  is the frequency.

### 2.3. Case Study

The present study was performed using annual rainfall data from 52 meteorological stations in Badajoz, provided by the Spanish Meteorological Institute.

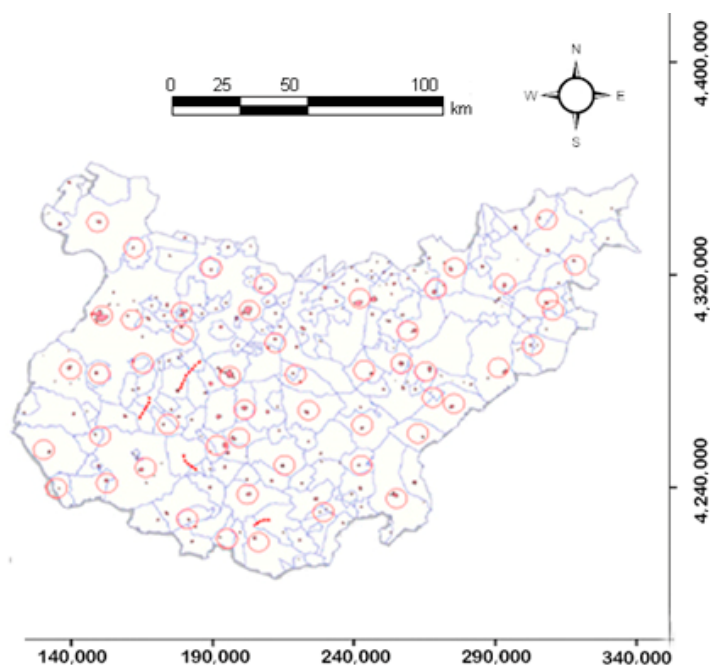
An exhaustive study was made using the statistical goodness-of-fit tests of Anderson–Darling and Kolmogorov–Smirnov to establish the degree of fitness of the Dagum distribution applied to the maximum annual rainfall series, and thus, be able to compare the fits of this distribution to those of the classical statistical distributions such as Gumbel, SQRT-ET max, Log-Pearson type III and the GEV.

Adjustment tests are widely used in hydrology due to the high degree of precision they provide when reflecting the fit of the statistical distribution to the rainfall data series available.

For the fit of the distributions, the software Easyfit was used. On introducing the maximum annual rainfall data, the program gives the degree of fitness of each statistical distribution according to the two goodness-of-fit tests used.

The 52 meteorological stations used in the analysis are shown in Table 1 and Figure 2. These contained maximum and minimum temperatures and daily precipitation for the period between 1990 and 2015. The quality control procedures of the Algorithm Theoretical Basis Document (ATBD) project, developed by the Royal Netherlands Meteorological Institute (KNMI) for the European Climate Assessment & Dataset (ECA&D), have been applied [43]. The blended series passed the standard homogeneity test, the Buishand range test, the Pettitt test and the Von Neumann ratio, as described by Wijngaard et al. [44] and ECA&D. Some series presenting missing values were completed following the recommendations of WMO [45] and Allen et al. [46]. The daily data from each station were processed and analyzed. Coefficients of variation and maximum precipitations at each meteorological station are also shown in Table 1.

The rainfall data from the 52 meteorological stations were introduced in EasyFit statistical program. The function of density was applied to the rainfall histogram for its adjustment. Finally, the goodness-of-fit to the rainfall histograms was studied as a function of the density of each statistical distribution.



**Figure 2.** Map of the spatial distribution of the meteorological stations. “Coordinate system is ETRS89/UTM zone 30” N.



**Table 1.** Meteorological stations used in the analysis.

Nº	Meteorological Stations	CV (%)	Pmax (mm)	Nº	Meteorological Stations	CV (%)	Pmax (mm)
1	San Vicente Alcántara	19.6	112.6	27	GranjaTorrehermosa	25.9	116.5
2	Alburquerque	22.3	104.2	28	Maguilla	27.9	90.3
3	La Roca de la Sierra	27.8	140.6	29	Peraleda del Zaucejo	30.2	75.3
4	Badajoz (Sagrajas)	29.5	102.2	30	Puebla del Prior	27.6	181.8
5	Montijo (Institute)	32.5	135.8	31	Puebla de la Reina	22.3	173.1
6	Olivenza	30.5	126.3	32	Monterrubio	19.6	98.7
7	Cheles	19.7	79.1	33	Puerto Hurraco	22.6	106.2
8	Alconchel	25.9	121.3	34	Castuera	27.4	95.4
9	Villanueva del Fresno	22.2	116.2	35	Quintana de Serena	25.4	102.3
10	Valencia del Mombuey	19.9	188.6	36	Valle de la Serena	29.6	89.7
11	Zahinos	30.1	163.3	37	Alange	22.8	119.5
12	Higuera de Vargas	25.6	217.4	38	Manchita	23.6	109.3
13	Jerez de los Caballeros	26.8	171.4	39	Guareña	28.4	116.5
14	Barcarrota	27.5	157.1	40	Aceuchal	26.8	92.5
15	La Albuera	31.2	146.6	41	Mérida	30.9	107.8
16	Talavera la Real	29.8	116.5	42	Santa Amalia	26.1	143.1
17	Fregenal de la Sierra	19.8	108.2	43	La Coronada	20.9	135.4
18	Segura de León	18.5	182.1	44	Campanario	19.9	99.1
19	Cabeza la Vaca	16.9	164.9	45	Acedera	22.6	76.8
20	Fuente de Cantos	19.6	124.5	46	Orellana la Sierra	28.5	148.3
21	Puebla del Maestre	28.6	134.2	47	Casas de Don Pedro	30.2	92.2
22	Casas de Reina	25.8	149.8	48	Capilla/Baterno	29.4	122.3
23	Villagarcía de la Torre	27.6	162.8	49	Siruella	25.4	145.5
24	Berlanga	28.9	134.9	50	Herrera del Duque	27.2	130.3
25	Valverde de Lereña	30.2	141.6	51	Villarta de Montes	22.1	140.1
26	Azuaga	27.4	150.6	52	Helechosa	25.6	98.5



### 3. Results

Figure 3 shows how the Dagum probability distribution fits the rainfall histogram in the town of San Vicente de Alcántara.

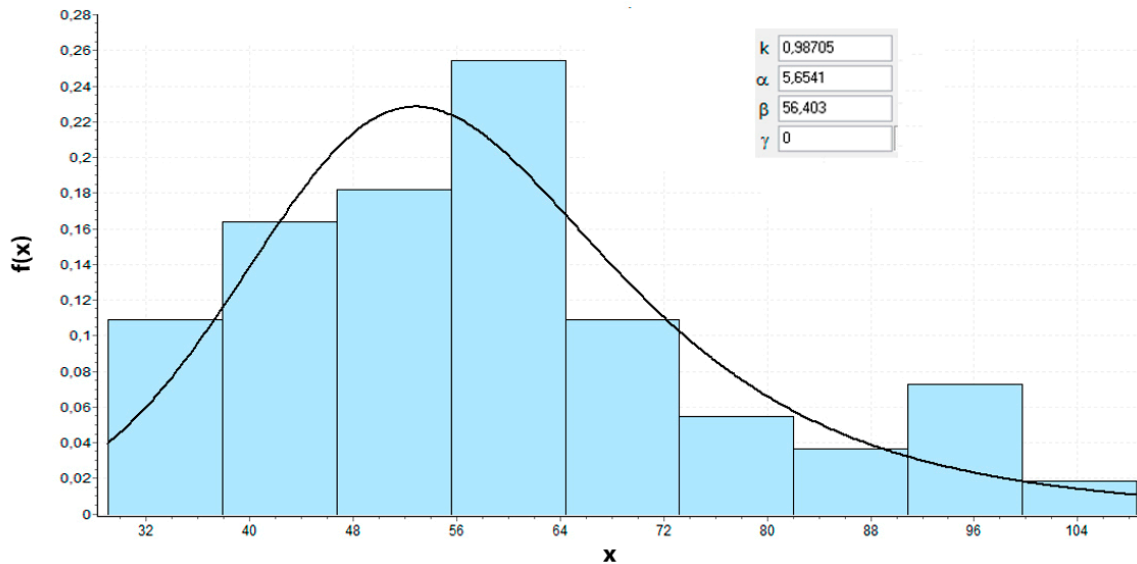


Figure 3. Dagum probability density distribution function in San Vicente de Alcántara for  $\gamma = 0$ .

Figure 4 shows how the density function fits the cumulative histogram of a maximum annual rainfall series. The density function of the statistical distribution is never going to reproduce the exact values of the histogram as desired that is why the goodness-to-fit tests are used to check which one provides the best fit to the rainfall data series when comparing various statistical distributions. That is to say, the chosen statistical distribution should be the one that fits the rainfall histogram more accurately.

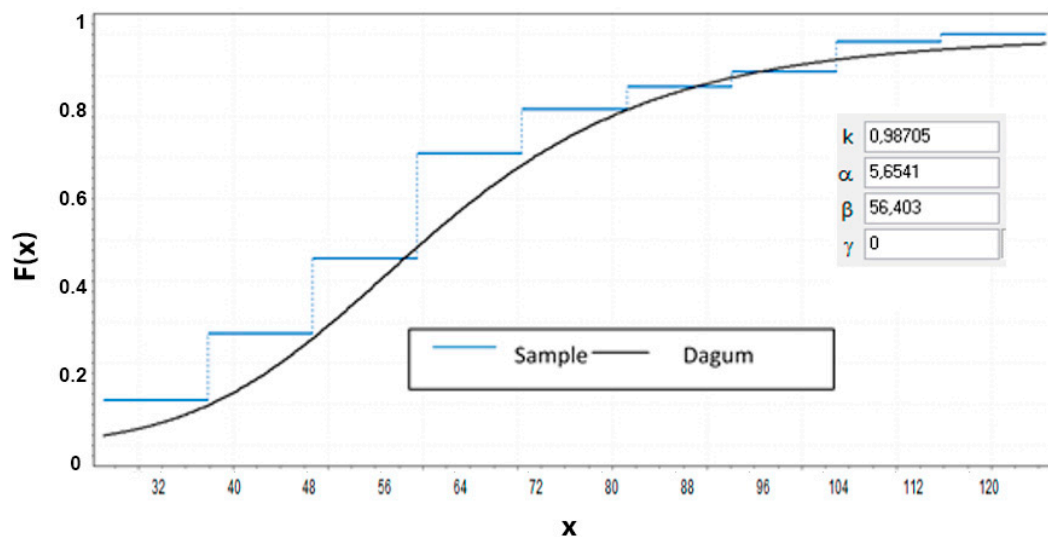
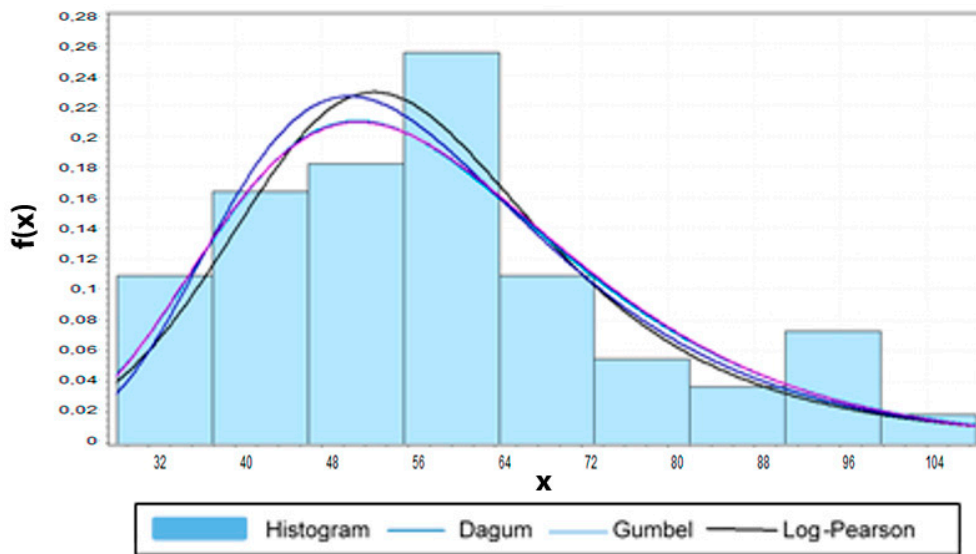


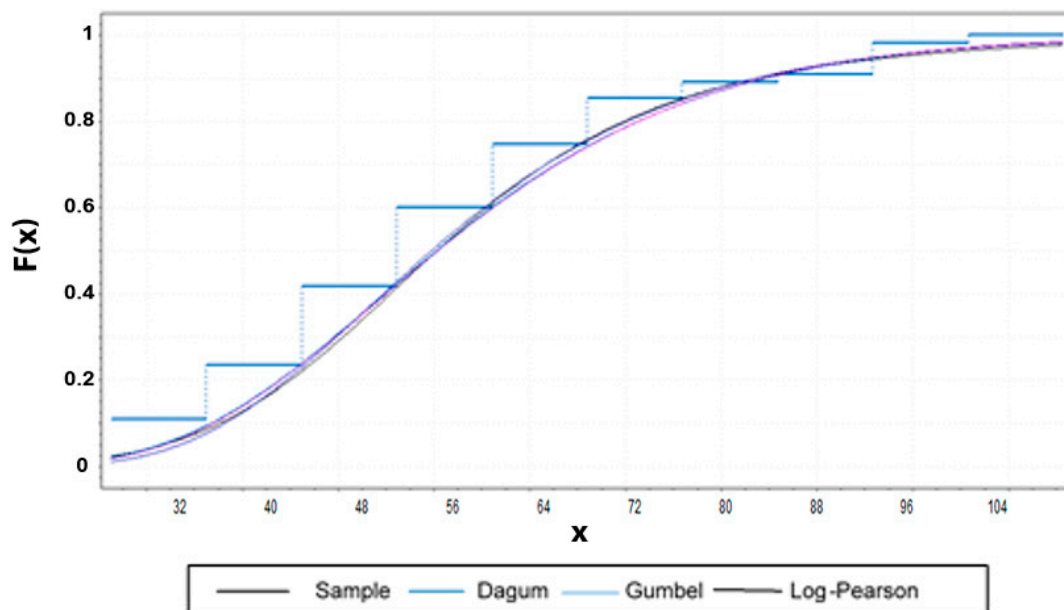
Figure 4. Cumulative distribution function for the Dagum distribution applicable to data from San Vicente de Alcántara.

After fitting the statistical functions to the rainfall data, Figure 5 shows graphically how the Dagum distribution is aligned with both the fit of Gumbel distribution and Log-Pearson type III distribution, being difficult to decide the best, since the curves are very close together.



**Figure 5.** Probability density function of Gumbel, Log-Pearson type III and Dagum Distributions. Comparison of the fit among the three different distributions applied to the histogram of the data from San Vicente de Alcántara.

Similarly, Figure 6 shows that it is difficult to determine which of the three distributions from all the cumulative distribution function fits more accurately the histogram of rainfall data in San Vicente de Alcántara.



**Figure 6.** Cumulative distribution functions of Gumbel, Log-Pearson type III and Dagum distributions, applied to the histogram of data from San Vicente de Alcántara.

As previously noted, it is difficult to draw any conclusions. Therefore, it is necessary to apply the goodness-to-fit tests. In this study, we applied the test to four distributions, as shown in Table 2. The distribution of Dagum clearly presents the lowest goodness-to-fit statistics in the San Vicente de Alcántara, Jerez de los Caballeros and Herrera del Duque data, which means that it fits better than the other three to the rainfall data.

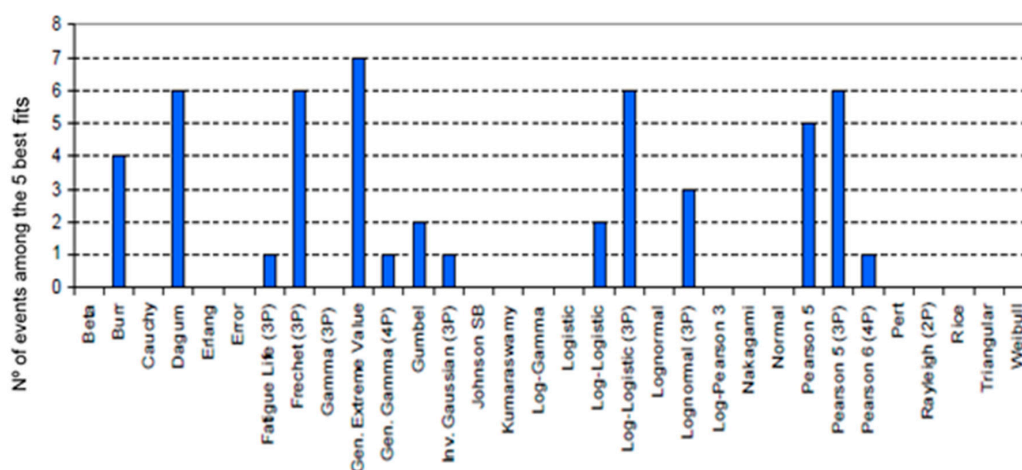
**Table 2.** Goodness-of-fit test results.

Distribution	San Vicente Alcántara		Jerez de los Caballeros		Herrera del Duque	
	Tests of Anderson–Darling	Test of Kolmogorov–Smirnov	Tests of Anderson–Darling	Test of Kolmogorov–Smirnov	Tests of Anderson–Darling	Test of Kolmogorov–Smirnov
Dagum	0.0490	0.1746	0.0421	0.0914	0.0391	0.1257
Log-Pearson III	0.0584	0.1947	0.0624	0.1758	0.5958	0.1985
Gumbel	0.0603	0.2553	0.0587	0.1404	0.0687	0.0984
GEV	0.0564	0.1875	0.0590	0.0654	0.0593	0.1751

These tests were applied to rainfall data from the remaining 51 stations [47], in which the analysis reflects a similar trend, and the statistical distribution Dagum presents lower goodness statistics than the rest of the distributions. Other tests were taken into account to select the best distribution model, such as the corrected Akaike information criterion and the Bayesian information criterion (BIC), but finally the tests of Anderson–Darling and the test of Kolmogorov–Smirnov were chosen.

Subsequently, and to confirm the above results, tests of goodness-of-fit were carried out with a series of statistical distributives (applied both in hydrology and in other disciplines), using the ten stations with the largest sample size (among the 52 stations).

In Figure 7, it can see that Dagum appears as one of the most frequent distributions (within the five best settings), just below the GEV distribution, but above the Gumbel and the Log-Pearson type III distribution.



**Figure 7.** Number of events in the different distributions according to the goodness-of-fit test at the locations for the ten largest simple sizes.

Table 3 shows that the application of the Dagum Distribution to the cases of Cabeza la Vaca, Monterrubio and Campanario obtain different flows to those of the traditional distributions. For example, Cabeza la Vaca shows 21.92% flow greater than the most commonly used distribution (Gumbel).

It can be deduced that the statistical distributions that provide the greatest rainfall are the Log-Logistic 3P distribution followed by the Dagum distribution. The quantitative differences in relation to the value provided by the Gumbel distribution, the most widespread in studies and projects, are variable and in some cases are even considerable, up to 58% higher in the case of Jerez de los Caballeros. Therefore, in order to be on the safety side, the Log-Logistic 3P distributions and the Dagum distribution, must be used, since you can be completely sure that the flows and precipitations derived from its application will be greater than the results obtained with the distributions of Gumbel, SQRT-ET max and Log Pearson 3. It is important to emphasize that these distributions are the best fit in the Kolmogorov–Smirnov, Anderson–Darling kindness tests and their weighting.

**Table 3.** Maximum daily precipitation for different distributions.

Distribution	Cabeza la Vaca		Monterrubio		Campanario	
	P24T(mm)	Variation Respect Gumbel	P24T(mm)	Variation Respect Gumbel	P24T(mm)	Variation Respect Gumbel
GEV	158.47	0.43%	98.23	−12.96%	89.03	−0.15%
Dagum	192.37	21.92%	123.04	9.28%	112.77	26.47%
Log-Logistic 3P	188.97	19.76%	124.04	9.90%	116.17	30.29%
Fechet 3P	151.64	−3.90%	110.94	−1.69%	85.35	−4.28%
Pearson 5 3P	146.19	−7.35%	100.38	11.05%	86.47	−3.02%
Gumbel	157.79	-	112.85	-	89.16	-
SQRT-ET max	164.90	4.51%	118.80	5.27%	93.45	4.81%
Log Pearson Type III	145.01	−8.10%	97.79	−13.34%	84.97	−4.71%
Regional analysis	172.00	9.00%	118.00	4.56%	98.00	9.91%

#### 4. Discussion

It is important to denote that the comparison of distributions is complicated. Goodness of fit tests are not very powerful and with the typical sample sizes available in practice it is rarely possible to reject statistically some distribution candidates. Therefore, the comparison must be done on a larger scale [48]. Thus, in this case study, sites located throughout the province considered are sufficiently numerous and evenly distributed to obtain significant results. However, these results cannot be extended elsewhere, that is, the choice of a particular distribution at a given place should be carefully studied and selected. Although some distributions, such as the Gumbel or Log-Pearson type III, have been extensively used in many hydrologic studies and without any additional consideration related to particular conditions of the basins, the inclusion of the spatial factor would reduce the uncertainty concerning the choice of the model [49]. Physical factors such as large-scale meteorological phenomena could create regional probabilities dependencies which have to be accounted for. In consequence, as it was previously indicated, each region or zone should be initially characterized for the choice of the statistical distribution, which better explains the expected rainfall events [50].

As a result of the analyses carried out in the 52 locations throughout the province of Badajoz, the Dagum model scored better than the other models which have been traditionally used in hydrologic and hydraulic works. There are few previous studies where the Dagum distribution had been used for these topics.

The Dagum model was found to overestimate a great number of times when compared to the Gumbel distribution. Therefore, the Dagum distribution seems to be the most recommendable distribution for a conservative design and for to plan accordingly [51].

Because of the ample availability of computers nowadays, many statistical distributions have to be considered when a single-site flood frequency analysis is done [52]. Moreover, as more data are being accumulated since the recent and coming years are providing more new information, new analyses could be performed with regionalized parameters of proven model for each location [53]. In this sense, the consideration of the Dagum model can provide more accurate results in many places of southwestern Spain that those obtained using traditional distributions. The evaluation and simulation of rainfall scenarios indicate that changes in rainfall characteristics have a considerable impact on the built drainage system and that Low Impact Development (LID) practices can adequately control flooding [54].

Future work should aim at verifying the applicability of the Dagum distribution in other regions of southern Europe.

#### 5. Conclusions

On analyzing the maximum annual rainfall data from 52 stations (strategically located throughout the zone) and treating them by using the goodness-of-fit tests of Anderson–Darling

and Kolmogorov–Smirnov, it is confirmed that, in addition to the distributions traditionally used in hydrology (such as the Gumbel, Log-Pearson type III and the EVD distribution), there is another statistical distribution, the Dagum, which can be used in hydrology and meets the formulation of extreme values (outliers) and fits better to the rainfall histograms.

Based on the statistical data from the study, it is concluded that the Dagum distribution presents lower statistics in the two goodness-of-fit tests mentioned above and, therefore, adjusts significantly better to the histograms of the maximum annual rainfall data than the commonly used distributions. Particularly, this new statistical distribution is more appropriate to reflect the rainfall regime in Badajoz.

In conclusion, the Dagum statistical distribution is proposed to improve hydrological studies in Badajoz, since the rainfalls given by its density function are more precise (as shown by the goodness-of-fit tests) than the rainfall data calculated through classical statistical distributions. Its use in the professional field would allow for greater flows rates to be considered when designing drainage systems and studies of flooding, thus preventing future possible rainfall damage.

**Author Contributions:** Conceptualization, F.L.-R.; Data curation, A.G.-C.; Formal analysis, F.M.-G.; Funding acquisition, F.M.-G.; Investigation, F.L.-R., J.G.-S.-C. and A.G.-C.; Methodology, F.L.-R.; Project administration, J.G.-S.-C.; Resources, F.M.-G.; Software, A.G.-C.; Supervision, J.G.-S.-C.; Validation, J.G.-S.-C.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors wish to acknowledge to the Junta de Extremadura and the European Social Found (FEDER) for the support of this research work. This study has been carried out through the Research Project GR-18029, GR-18081 and GR-18086 linked to the VI Regional Plan for Research, Technological Development and Innovation from the General Government of Extremadura 2017–2020.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Sanches Fernandes, L.; Pereira, M.-G.; Morgado, S.G.; Macário, E.B. Influence of Climate Change on the Design of Retention Basins in Northeastern Portugal. *Water* **2018**, *10*, 743. [[CrossRef](#)]
2. Bai, T.; Mayer, A.L.; Shuster, W.D.; Tian, G. The Hydrologic Role of Urban Green Space in Mitigating Flooding (Luohe, China). *Sustainability* **2018**, *10*, 3584. [[CrossRef](#)]
3. Gumbel, E.J. *Multivariate Extremal Distributions*; Bull. Inst. International the Statistique: Paris, France, 1960.
4. Chow, V.T.D.R.; Maidment, L.W.M. *Hidrología Aplicada*; McGraw-Hill: Colombia, Santa Fé de Bogotá, 1994.
5. Ponce, V.M. *Engineering Hydrology, Principles and Practices*; Prentice Hall: Upper Saddle River, NJ, USA, 1989.
6. Pearson, K. *Mathematical Contributions to the Theory of Evolution, XIX: Second Supplement to a Memoir on Skew Variation*; Philosophical Transactions of the Royal Society: London, UK, 1916.
7. Ferrer, F.J. *Obtención de la lluvia neta según la metodología del Soil Conservation Service*; Cedex: Madrid, Spain, 1991. (In Spanish)
8. Témez, J.R. *Extended and Improved Rational Method. Version of the Highways Administration of Spain*; XXIV Congreso Internacional de la IAHR: Madrid, Spain, 1991.
9. Kotz, S.; Nadarajah, S. *Extreme Value Distributions: Theory and Applications*; Imperial College Press: London, UK, 2000.
10. Zorraquino, C. *La Función SQRT-ET Max*; Revista de Obras Públicas: Madrid, Spain, 2004. (In Spanish)
11. de Fomento, M. *Máximas Lluvias diarias en la España Peninsular*; Ministerio de Fomento: Madrid, Spain, 1999. (In Spanish)
12. Ministerio de Obras Públicas y Urbanismo. *Instrucción de Carreteras, Norma 5.2-IC de Drenaje Superficial. It has been doneMadrid*; It has been doneMadrid; Ministerio de Obras Públicas y Urbanismo: Madrid, Spain, 1990. (In Spanish)
13. Catalá, F. *Cálculo de caudales en las redes de saneamiento*; Colegio de Ingenieros de Caminos, Canales y Puertos: Madrid, Spain, 1997.
14. Ferrer, F.J. *Recomendaciones para el cálculo hidrometeorológico de avenidas*; Cedex: Madrid, Spain, 1993. (In Spanish)

15. Panagoulia, D.; Caroni, C. Modelling maximum precipitation in a mountainous area of Greece under global warming, Hydro-climatology: Variability and Change, Proceedings of Symposium J-H02 held during IUGG2011 in Melbourne, Australia. *IAHS Publ.* **2011**, *344*, 176–181.
16. Lang, M.; Ouarda, T.B.M.J.; Bobée, B. Towards operational guidelines for over-threshold modeling. *J. Hydrol.* **1999**, *225*, 103–117. [[CrossRef](#)]
17. Onyutha, C. On rigorous drought assessment using daily time scale: Non-stationary frequency analyses, revisited concepts, and a new method to yield non-parametric indices. *Hydrology* **2017**, *4*, 48. [[CrossRef](#)]
18. Onyutha, C. Hydrological model supported by a step-Wise calibration against sub-flows and validation of extreme flow events. *Water* **2019**, *11*, 244. [[CrossRef](#)]
19. Pickands, J. Statistical inference using extreme order statistics. *Ann. Stat.* **1975**, *3*, 119–131.
20. Modarres, R.; Sarhadi, A. Statistically-based regionalization of rainfall climates of Iran. *Glob. Planet. Change* **2011**, *75*, 67–75. [[CrossRef](#)]
21. Dagum, C. Un nuevo modelo de distribución del ingreso personal: Especificación y estimación. *Economía Aplicada* **1977**, *30*, 413–437. (In Spanish)
22. Kleiber, C. A guide to the Dagum distributions. In *Modeling Income Distributions and Lorenz Curves*; Springer: New York, NY, USA, 2008; pp. 97–117.
23. Mielke, P.R., Jr.; Johnson, E.S. Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resour. Res.* **1974**, *10*, 223–226. [[CrossRef](#)]
24. Alam, S.; Sabbir, M.; Khan, M. Statistical Characterization of Extreme Hydrologic Parameters for the Peripheral River System of Dhaka City. *J. Water Resour. Ocean Sci.* **2014**, *3*, 30–37. [[CrossRef](#)]
25. Alam, S.; Khan, S.M.; Rahat, S.H. A Study on Selection of Probability Distributions of Extreme Hydrologic Parameters for the Peripheral River System of Dhaka City. In Proceedings of the Academics World 15th International Conference, Bangkok, Thailand, 28–29 December 2015; pp. 29–34.
26. Crabbe, P.J. Conceptual Framework for a Systems Dynamics Adaptation Model to Climate Change for Charlottetown, P.E.I., Canada. *J. Glob. Resour.* **2016**, *2*, 1–11.
27. Mayooraan, T.; Laheetharan, A. The Statistical Distribution of Annual Maximum Rainfall in Colombo District. *Sri Lankan J. Appl. Stat.* **2014**, *15*, 107–130. [[CrossRef](#)]
28. Pérez-Sánchez, J.; Senent-Aparicio, J. Analysis of meteorological droughts and dry spells in semiarid regions: A comparative analysis of probability distribution functions in the Segura Basin (SE Spain). *Theor. Appl. Climatol.* **2018**, *133*, 1061. [[CrossRef](#)]
29. Domma, F.; Condino, F. The Beta-Dagum distribution: Definition and properties. *Commun. Stat. Theory Methods* **2013**, *42*, 4070–4090. [[CrossRef](#)]
30. Cheng, Y.; Wang, R. A novel stormwater management system for urban roads in China based on local conditions. *Sustain. Cities Soc.* **2018**, *39*, 163–171. [[CrossRef](#)]
31. Douka, M.; Karacostas, T. Statistical analyses of extreme rainfall events in Thessaloniki, Greece. *Atmos. Res.* **2018**, *208*, 60–77. [[CrossRef](#)]
32. Shahzad, M.N.; Asghar, Z. Transmuted dagum distribution: A more flexible and broad shaped hazard function model. *Hacet. J. Math. Stat.* **2016**, *45*, 227–244. [[CrossRef](#)]
33. Panagoulia, D.; Economou, P.; Caroni, C. Stationary and non-stationary generalized extreme value modeling of extreme precipitation over a mountainous area under climate change. *Environmetrics* **2014**, *25*, 29–43. [[CrossRef](#)]
34. Caroni, C.; Panagoulia, D.; Economou, P. Non-stationary modelling of extremes of precipitation and temperature over mountainous areas under climate change. In Proceedings of the International Conference in Current Topics on Risk Analysis: ICRA6 and RISK, Barcelona, Spain, 26–29 May 2015; pp. 203–209.
35. Anderson, T.W.; Darling, D.A. Asymptotic theory of certain “goodness-of-fit” criteria based on stochastic processes. *Ann. Math. Stat.* **1952**, *23*, 193–212. [[CrossRef](#)]
36. Gumbel, E.J. Les valeurs extrêmes des distributions statistiques. *Ann. Inst. Henri Poincaré* **1935**, *5*, 115–158.
37. Champernowne, D.G. The graduation of income distributions. *Econ. J. Econom. Soc.* **1952**, *20*, 591–615. [[CrossRef](#)]
38. MATHWAVE. Available online: <http://www.mathwave.com/easyfit-distribution-fitting.html> (accessed on 15 December 2018).
39. Kolmogorov, A.N. *Foundations of the Theory of Probability*; Chelsea Publishing Company: New York, NY, USA, 1950.



40. Miller, I.; Freund, J.E. *Probabilidad y estadística para ingenieros*; Reverté: Madrid, Spain, 1969.
41. Shorack, G.R.; Wellner, J.A. *Empirical Processes with Applications to Statistics*; Wiley: New York, NY, USA, 1986.
42. Anderson, T.W.; Darling, D.A. A Test of Goodness-of-Fit. *J. Am. Stat. Assoc.* **1954**, *49*, 765–769. [[CrossRef](#)]
43. ECA&D. *Algorithm Basis Document (ATBD)*; Royal Netherlands Meteorological Institute KNMI: De Bilt, The Netherlands, 2013.
44. Wijngaard, J.B.; Tank, A.K.; Können, G.P. Homogeneity of 20th century European daily temperature and precipitation series. *Int. J. Climatol.* **2003**, *23*, 679–692. [[CrossRef](#)]
45. WMO. *Guide to Climatological Practices (WMO No. 100)*, 3rd ed.; WMO: Geneva, Switzerland, 2011.
46. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements*; FAO Irrigation and Drainage Paper, N° 56; FAO: Rome, Italy, 1998.
47. García-Conde, A.J. Análisis de distribuciones estadísticas alternativas a las tradicionales para la optimización de los caudales de cálculo empleados en los estudios de hidrología. Ph.D. Thesis, University of Extremadura, Badajoz, Spain, 2013. (In Spanish)
48. Bobée, B.; Cavadis, G.; Ashkar, F.; Bernier, J.; Rasmussen, P. Towards a systematic approach to comparing distributions used in flood frequency analysis. *J. Hidrol.* **1993**, *142*, 121–136. [[CrossRef](#)]
49. Haktanir, T.; Horlacher, H.B. Evaluation of various distributions for flood frequency analysis. *Hidrol. Sci. J. Sci. Hydrol.* **1993**, *38*, 15–32. [[CrossRef](#)]
50. Courty, L.G.; Rico-Ramirez, M.A.; Pedrozo-Acuña, A. The Significance of the Spatial Variability of Rainfall on the Numerical Simulation of Urban Floods. *Water* **2018**, *10*, 207. [[CrossRef](#)]
51. Candelario-Garrido, A.; García-Sanz-Calcedo, J.; Reyes Rodríguez, A.M. A quantitative analysis on the feasibility of 4D planning graphic systems versus conventional systems in building projects. *Sustain. Cities Soc.* **2017**, *35*, 378–384. [[CrossRef](#)]
52. Panagoulia, D.; Economou, P.; Caroni, C. Quantifying uncertainty in modelled estimates of annual maximum precipitation: Confidence intervals, European Geosciences Union. Precipitation uncertainty and variability: Observations, ensemble simulation and downscaling. *Geophys. Res. Abstr.* **2016**, *18*, 17–22.
53. Carretero-Ayuso, M.J.; García-Sanz-Calcedo, J. Analytical study on design deficiencies in the envelope projects of healthcare buildings in Spain. *Sustain. Cities Soc.* **2018**, *42*, 139–147. [[CrossRef](#)]
54. Zhihua, Z.; Xiaohong, C. Evaluating the Effects of Low Impact Development Practices on Urban Flooding under Different Rainfall Intensities. *Water* **2017**, *9*, 548. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).