

## Best-fit probability distributions and return periods for daily rainfall in the Kyrenia region, Northern Cyprus

**Youssef Kassem**<sup>a1</sup>, Near East University, Department of Mechanical Engineering, Engineering Faculty, 99138 Nicosia (via Mersin 10, Turkey), Cyprus.

**Hüseyin Gökçekuş**<sup>b</sup>, Near East University, Civil and Environmental Engineering Faculty, Department of Civil Engineering, 99138 Nicosia (via Mersin 10, Turkey), Cyprus

**Nour Alijl**<sup>c</sup>, Near East University, Civil and Environmental Engineering Faculty, Department of Civil Engineering, 99138 Nicosia (via Mersin 10, Turkey), Cyprus.

### Suggested Citation:

Kassem, Y., Gökçekuş, H. & Alijl, N. (2021). Best-fit probability distributions and return periods for daily rainfall in the Kyrenia region, Northern Cyprus. *World Journal of Environmental Research*. 11(2), 39-48. <https://doi.org/10.18844/wjer.v11i2.7189>

Received from August 11, 2021; revised from October 15, 2021; accepted from December 01, 2021; Selection and peer-review under responsibility of Murat Sonmez, Middle East Technical University, Northern Cyprus Campus, Cyprus

©2021 Birlesik Dunya Yenilik Arastirma ve Yayıncılık Merkezi. All rights reserved.

### Abstract

Kyrenia region is in the northern part of Northern Cyprus that is environmentally fragile and susceptible to natural disasters. Thus, the study of frequency analysis is essential to find the most suitable model that could detect the region's risk in certain natural phenomena such as rainfall, flood, and so on. The objective of this research is to determine the best fit probability distribution in the case of average daily rainfall and total rainfall using 22 years of data (1995-2016) from the Kyrenia region in Northern Cyprus by using 37 probability distribution models. The best-fit probability distribution in the case of maximum annual daily rainfall is determined using various distribution types. Three goodness-of-fit test statistics were applied. Beta, Dagum, Wakeby, Paretoa, Log-Pearson 3, Gen. Extreme Value, and Gen. Gamma (4P) showed the largest number of best-fit results. The results of this study can be used to develop more accurate models of flooding risk and damage.

**Keywords:** Distribution function; goodness-of-fit tests; Northern Cyprus; rainfall.

\* ADDRESS FOR CORRESPONDENCE: Youssef Kassem, Near East University, Department of Mechanical Engineering, Engineering Faculty, 99138 Nicosia (via Mersin 10, Turkey), Cyprus.  
E-mail address: [yousseuf.kassem@neu.edu.tr](mailto:yousseuf.kassem@neu.edu.tr)

## 1. Introduction

The amount and pattern of the rainfall in a specific region are essential factors that affect various natural and socio-economic systems such as flood protection, water resources management, agriculture and forestry, and tourism (Alam, Emura, Farnham & Yuan, 2018). Thus, analysis of rainfall is a basic tool for urban infrastructure planning and design such as culverts and urban drainage systems (Mamoon & Rahman, 2016). It depends on the characteristics of the rainfall pattern (Meena, Dubey & Basak, 2019). In general, rainfall is an important input to rainfall-runoff modeling. In hydrological design applications, design rainfall (intensity–duration–frequency) data is used to obtain rainfall intensity at a given region for a given duration (Gratien et al., 2019). The intensity duration frequency data are widely used in the planning and designing of stormwater infrastructure and flood management works (Tfwala et al., 2017). In deriving intensity duration frequency curves, one of the primary steps is fitting an appropriate probability distribution to at-site rainfall data. Additionally, according to Michaelides et al. (2009), no theoretical distribution can be considered that can characterize exclusively the rainfall profile. Therefore, the selection of a probability distribution that gives the best fit to the observed rainfall or flood data is an important research topic in the field of statistical hydrology (Okoli, Mazzoleni, Breinl & Di Baldassarre, 2019).

Numerous scientific studies have been conducted worldwide on the selection of probability distributions in rainfall frequency analyses. For instance, Yuan et al. (2018) utilized the Expanded Automated Meteorological Data Acquisition System (EA) weather data for 20 years (1981-2000) to analyze the annual maximum hourly rainfall characteristics for 15 locations in Japan. The results showed that the Log-Pearson type 3 distributions provided the best fit to the actual data for most locations in Japan. Parchure and Gedam (2019) determined the best-fit probability distribution for extreme storms using two-parameter and three-parameter distribution functions. The results indicated that a three-parameter generalized extreme value was considered the best distribution function to study the extreme storm series in the Mumbai region, India. Kassem and Gökçekuş (2020) utilized four distribution functions to analyze the characteristics of rainfall in the Beirut region, Lebanon using daily rainfall of 25 years (1991–2015). The results showed that Gumbel Maximum and Logistic distributions were able to provide the best fit to the actual data for the selected region.

Cyprus is the third largest island in the Mediterranean Sea with an area of 9,251km<sup>2</sup>. It has a temperate climate (Mediterranean climate). Many of the highly populated regions in Cyprus, particularly the Northern part of Cyprus that is located on the coast, for example, Kyrenia (Girne), Gazimağusa (Famagusta), Güzelyurt (Morphou) are highly susceptible to urban flooding. One of the main reasons cited for this is rapid urbanization, which causes changes in landscape owing to the construction of urban infrastructures and changes in runoff conveyance networks. Kyrenia region, having an area of 690m<sup>2</sup> with a population of 20851, has been inflicted by negative impacts from flooding due to heavy and torrential rainfall in its urban environment. Many flash floods that occurred in the Kyrenia region were reported by news and social media (UNOOSA, 2010; Kratzer, 2014; FloodList, 2018; LGC News, 2020).

To the best of our knowledge, there are no detailed studies in the Kyrenia region regarding frequency analysis and estimating the daily rainfall as a function of the number of days and meteorological parameters. Additionally, according to the authors' review, there is no major study of frequency analysis on the comparison of probability distributions for finding the most suitable model that could anticipate extreme events of rainfall. Therefore, this study aims to study the characteristics of the distribution of rainfall in the Kyrenia region, Northern Cyprus. In the present study, 37 probability distribution models are used to determine the best fit probability distribution in the case of average daily rainfall, total rainfall, and maximum daily rainfall using 22 years of data (1995-2016). Goodness-of-fit tests including Kolmogorov–Smirnov (K-S) test, Anderson–Darling (A-D) test, and Chi-squared (C-s) test are used to select

the best fit probability and distribution model. Moreover, six formulas are used to expect the return period in years of daily maximum rainfall in the selected region.

### **1.1. Purpose of study**

The objective of this research is to determine the best fit probability distribution in the case of average daily rainfall and total rainfall using 22 years of data (1995-2016) from the Kyrenia region in Northern Cyprus by using 37 probability distribution models.

## **2. Materials and Method**

### **2.1. Study Area**

Kyrenia is located in the northwestern part of Cyprus at a latitude of 35°20'30.00" N, the longitude of 33°19'00.00" E, and surrounded by the five fingers mountain (Beşparmak), where elevations range from - 2 to 1026 m above sea level, which has major impact into defining the weather on the island and slopes range from 0 % to 20%. Rainfall distribution over the country varies considerably among regions. The minimum and maximum average values of rainfall are occurred in Mesaoria, with 300 to 400 mm, and the Karpas Peninsula area averages 400 to 450 mm of annual rainfall, respectively. Snow is rarely occurring on the upper hills of the Beşparmak Mountains in the North and is usually available at the peaks of the Troodos Mountains throughout the year in South Cyprus. The maximum snowfall over the Northern part of Cyprus in the last ten years (1992-2002) was measured to be 15 cm on Beşparmak, Selvili Tepe, and Katara hills. Additionally, temperatures are ranged from 21°C along the coast to 15°C on top of the Kyrenia Range. The maximum temperature can reach 35- 45°C in the Mesaoria Plain.

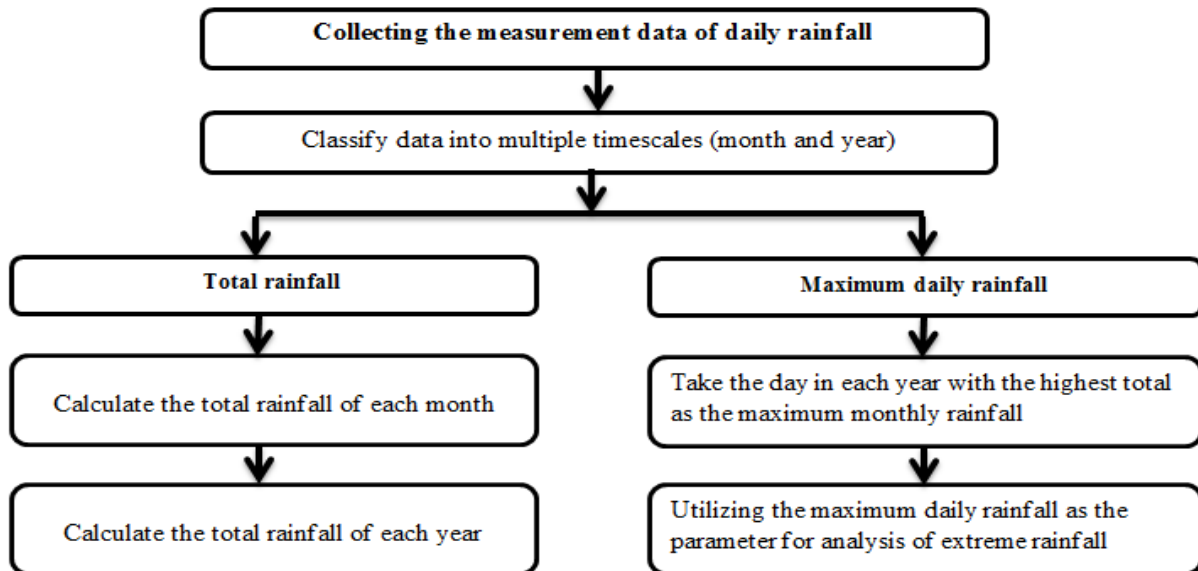
### **2.2. Data and Measurement**

The daily measurement data including rainfall, average temperature, minimum and maximum temperatures, solar radiation, and wind speed were collected from the Meteorological department located in North Nicosia (Lefkoşa) during 1995-2015. The data were measured at various heights. The rain gauge was used to measure the rainfall at a height of 0.3m above ground level.

#### **2.2.1. Probability Distributions**

The choice of the probability distribution models is important to select the best-fit probability distribution for a specific location. The objective of this study is to determine the best-fit probability distributions in the case of daily rainfall, annual rainfall, and maximum daily rainfall using 22 years of data (1995-2016) from the Kyrenia region in the Northern Part of Cyprus using various types of distribution. Figure 1 illustrates the procedure for calculating the total rainfall and maximum daily rainfall. In this section, 37 distributions models (Beta, Four-Parameter Burr, Three-Parameter Burr, Cauchy, Four-Parameter Dagum, Three-Parameter Dagum, Three-Parameter Erlang, Two-Parameter Erlang, Two-Parameter Exponential, One-Parameter Exponential, Three-Parameter Gamma, Two-Parameter Gamma, Generalized Extreme Value, Four-Parameter Generalized Gamma, Three-Parameter Generalized Gamma, Generalized Logistic, Generalized Pareto, Maximum Extreme Value Type 1, Minimum Extreme Value Type 1, Three-Parameter Inverse Gaussian, Log-Gamma, Logistic, Two-Parameter Inverse Gaussian, Log-Logistic, Three-Parameter Lognormal, Two-Parameter Lognormal, Log-Pearson 3, Nakagami, Normal, Two-Parameter Rayleigh, One-Parameter Rayleigh, Wakeby, Three-Parameter Weibull, Two-Parameter Weibull) were selected to analyze the characteristic of daily rainfall in the Kyrenia region in Northern Cyprus. The method of maximum-likelihood is utilized to estimate the parameters of distribution models.

**Figure 1**  
*Procedure for calculating the total rainfall and extreme rainfall*



### 2.3. Analysis

The estimation of the return period of high intensity of rainfall is very important for many engineering aspects including the design of contour drains, dams, road culverts, airfield drainage, storm sewers in urban areas, and flood or sediment control dams in water catchments. The return period is a measure of the probable time interval between the occurrence of a given event and that of an equal or greater event. If the variable ( $X$ ) equal to or greater than  $x$  occurs on the average once in  $T$  years, then the probability of occurrence  $P(X \geq x)$  of such a variable is shown in the below equation (Alam, Emura, Farnham, & Yuan, 2018)., 30].

$$\begin{aligned}
 P &= \frac{1}{T} (X \geq x) \quad \text{or} \quad T \\
 &= \frac{1}{P} (X \\
 &\geq x)
 \end{aligned}
 \tag{1}$$

To determine the plotting position, which refers to the probability value assigned to each piece of data to be plotted, several methods have been proposed. Equation (2) is the most plotting position formula represented by

$$\begin{aligned}
 P &= \frac{1}{T} \\
 &= \frac{m - b}{n + 1 - 2b}
 \end{aligned}
 \tag{2}$$

where  $m$  is the rank of a value in a list ordered by descending magnitude,  $n$  is the total number of values to be plotted,  $b$  is a parameter, which is different in different formulas ( $b = 0.5$  for Hazen;  $b = 0.3$  for Chegodayev;  $b = 0$  for Weibull;  $b = 3/88$  for Blom;  $b = 1/3$  for Tukey; and  $b = 0.44$  for Gringoten).

## 3. Results

### 3.1. Rainfall characteristics

In this section, the daily rainfall ( $R$ ) data are analyzed statistically. The statistical characteristics include arithmetic mean (Mean) standard deviation (SD), coefficient of percent variation (CV), minimum (Min.),

maximum (Max.), skewness (S), and kurtosis (K), of daily rainfall for the selected region, are summarized in Table 1. It is found that the mean values of daily rainfall are within the range of 0.501-1.624mm. The maximum value of daily rainfall occurred in December 2001 (02/12/2020) with a value of 40.14mm and the minimum value of 0mm was recorded in the summer season for whole years.

**Table 1**

*Statistical estimators of the mean daily rainfall for the period 1995-2016*

Year	Mean	SD	CV	Min.	Max.	S	K
1995	0.5840	1.8737	320.83	0.0000	17.0800	5.28	32.69
1996	0.968	2.714	280.31	0.000	30.570	5.65	45.68
1997	0.865	2.412	278.85	0.000	20.360	4.33	22.69
1998	0.836	2.448	292.75	0.000	21.050	4.52	24.23
1999	0.5175	1.6616	321.11	0.0000	18.5300	5.91	46.92
2000	0.946	3.038	321.12	0.000	39.600	7.44	78.48
2001	1.165	3.551	304.73	0.000	40.140	6.16	51.15
2002	0.824	2.421	293.73	0.000	25.620	5.34	38.61
2003	0.970	2.608	269.05	0.000	17.180	3.88	16.17
2004	1.107	3.443	311.07	0.000	24.030	4.13	18.46
2005	0.783	2.472	315.67	0.000	16.090	4.56	21.75
2006	0.787	2.555	324.73	0.000	29.880	6.62	58.43
2007	0.967	3.256	336.84	0.000	32.390	5.70	38.99
2008	0.629	2.404	382.12	0.000	22.370	5.84	39.28
2009	1.460	3.851	263.83	0.000	33.950	4.35	24.61
2010	0.907	3.148	347.05	0.000	30.670	5.55	37.58
2011	1.086	2.793	257.21	0.000	19.440	3.94	17.99
2012	1.624	4.078	251.12	0.000	29.140	3.60	15.28
2013	0.5008	1.2567	250.95	0.0000	8.6400	3.74	16.54
2014	0.974	2.813	288.83	0.000	29.040	5.05	34.48
2015	1.027	2.825	275.01	0.000	27.850	4.85	30.90
2016	0.871	3.073	352.75	0.000	39.430	7.35	74.36
Average	0.9275	0.9730	104.91	0.0005	4.0959	1.16	0.65
Total	19.72	21.11	107.04	0.01	89.93	1.22	0.84

### **3.2. Probability distribution function and selecting the best-fit result for average daily rainfall and the total amount of rainfall**

The distribution parameters were calculated using mean daily rainfall with the maximum likelihood method. The best distribution among the 37 distribution functions for the selected location was evaluated based on the Kolmogorov-Smirnov (K-S) test, Anderson-Darling (A-D) test, and Chi-squared (C-s) test. Generally, the distribution with the lowest K-S, A-D, and C-s values will be selected to be the best model for the rainfall distribution in the studied location. Additionally, Table 2 shows the results of the goodness-of-fit for the best-fit probability distribution. Based on the K-S test, the Beta distribution has the lowest value, which is considered the best distribution function to study the average daily rainfall characteristics. Based on the A-D and C-s tests, Dagum is among the distribution giving the best fits to investigate the average daily rainfall distribution in the selected regions. Also, it is observed that the inverse Gaussian and inverse Gaussian (3P) distribution functions cannot be used to investigate the average daily rainfall in the studied location based on goodness-of-fit tests, as shown in Table 2.

Furthermore, it is observed that Wakeby is considered the best model to study the total rainfall distribution based on the K-S and A-D tests as shown in Table 2. Based C-s test, Pareto is among the distribution giving the best fits to investigate the total rainfall distribution, respectively (Table 2). Figure 2 illustrates the frequency histograms and probability plots of rainfall in the selected region.

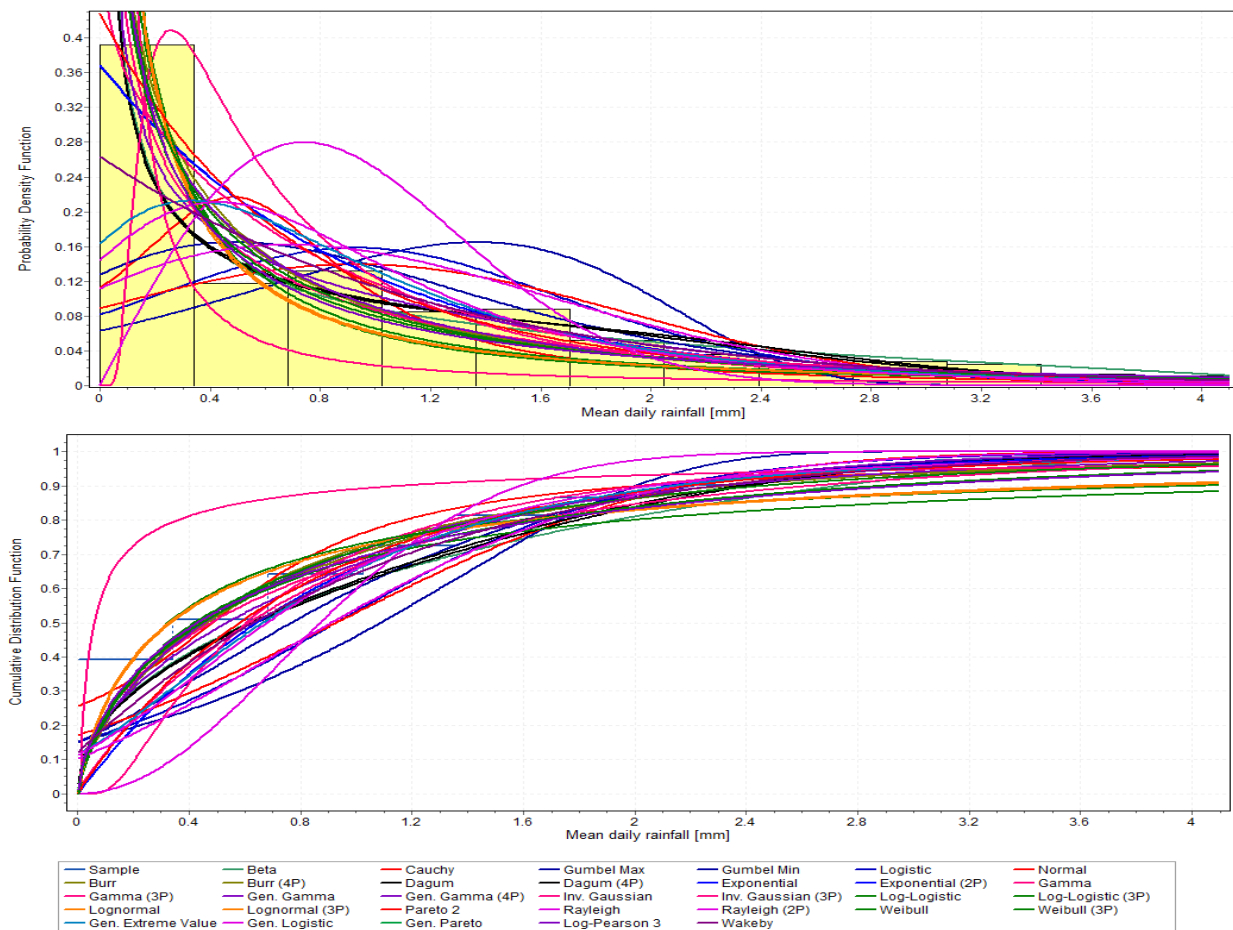
**Table 2**

*Results of goodness-of-fit of the best-fit distribution function (Rank # 1)*

Case	Distribution	Goodness-of-fit test		
		K-S	D-A	C-s
Average daily rainfall	Beta	0.0637		-
	Dagum	-	1.5748	-
	Dagum	-		11.825
Total rainfall	Wakeby	0.08771	-	-
	Wakeby	-	0.17379	-
	Pareto	-	-	0.09692

**Figure 2**

*Frequency histograms, probability density function, and cumulative distribution function plots of average daily rainfall (1995-2016)*



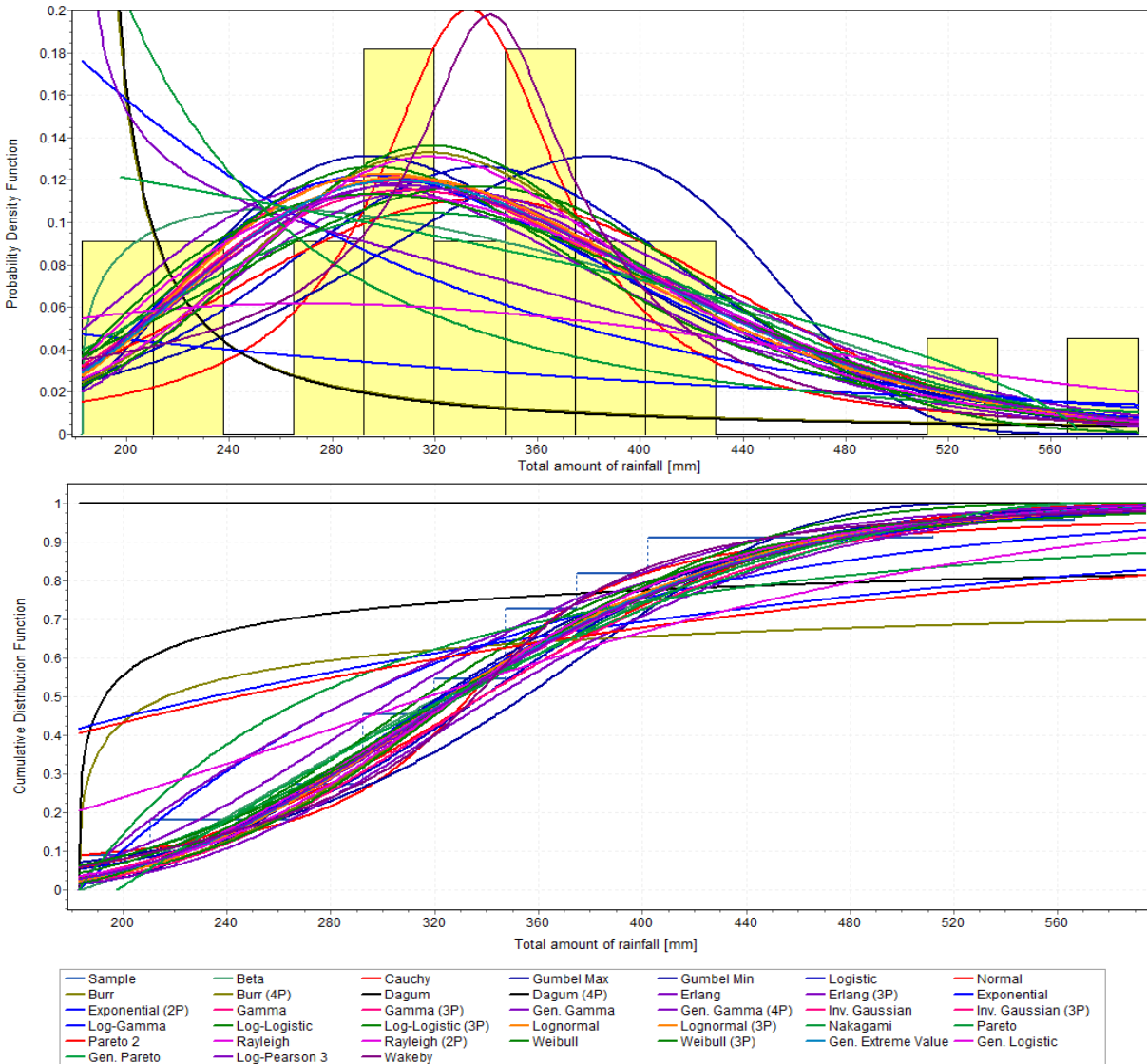
**3.3. Probability distribution function and selecting the best-fit result for extreme rainfall estimation**

In general, one of the primary natural causes of flooding is extreme rainfall events. Also, the ability to predict the extremes of rainfall would aid in improved flood planning in the specific region. Thus, frequency analysis is utilized to predict how often certain values of a variable phenomenon may occur and to assess the reliability of the prediction. In this study, 37 different probability distribution functions were used to

predict the probability distribution of the occurrence of annual maximum daily rainfall. Based on the K-S tests, Log-Pearson 3 distribution has the lowest value, which is considered the best distribution function to study the average rainfall characteristics (Table 3). Additionally, based on the A-D tests, Gen. Extreme Value is among the distribution giving the best fits to investigate the maximum daily rainfall distribution (Table 3).

**Figure 3**

*Frequency histograms, probability density function, and cumulative distribution function plots of average daily rainfall (1995-2016)*



Moreover, Gen. Gamma (4P) is the best overall model according to the C-s test for the selected location (Table 3). Also, it is observed that the Exponential, Pareto, and Pareto 2 distribution functions cannot be used to investigate the average rainfall in the studied location based on C-s and A-D tests.

**Table 3**

*Results of goodness-of-fit of the best-fit distribution function (Rank # 1)*

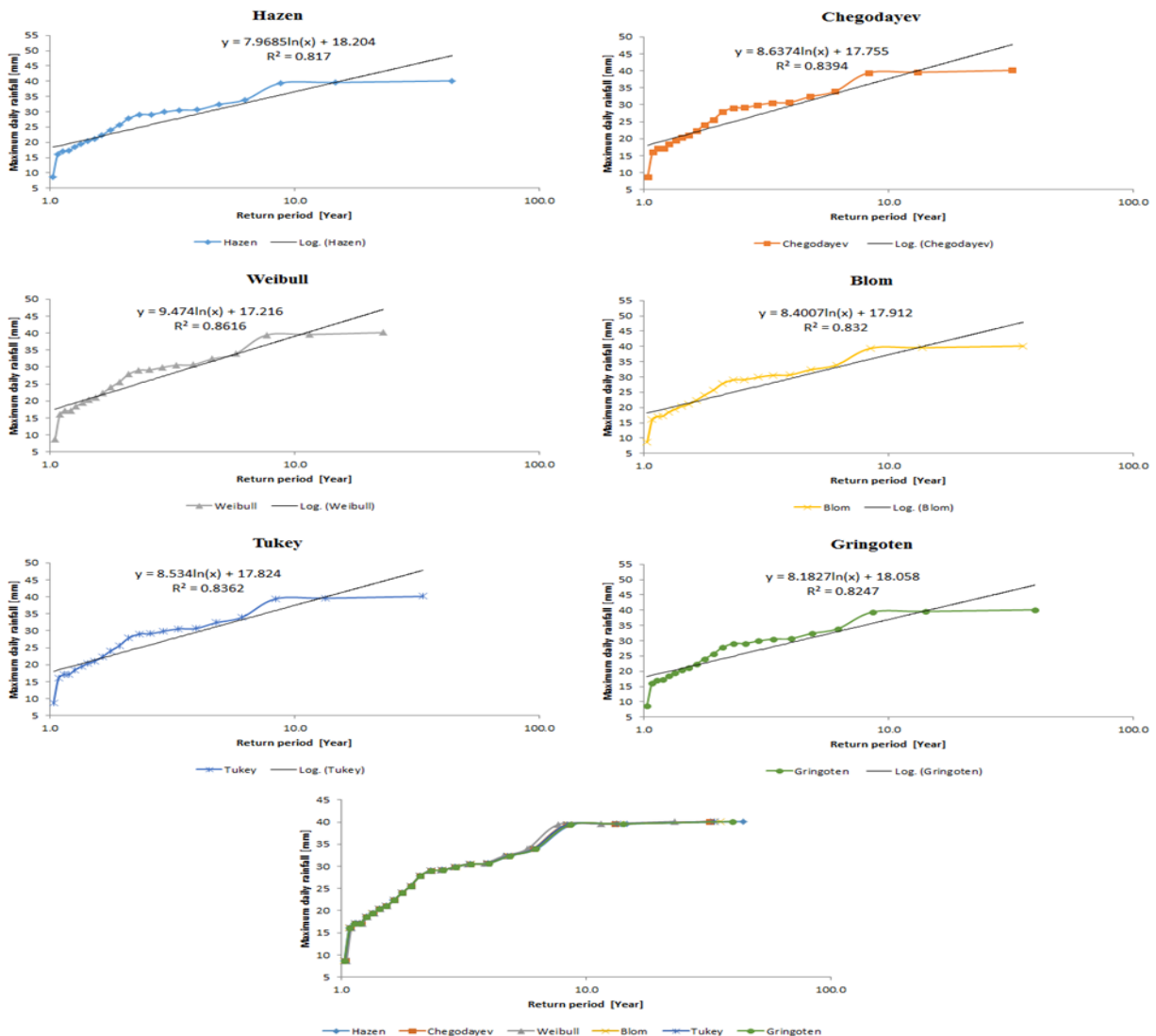
Distribution	K-S	Distribution	A-D	Distribution	C-s
Log-Pearson 3	0.08139	Gen. Extreme Value	0.24113	Gen. Gamma (4P)	0.03041

### 3.4. Return period

In this section, the return period of annual maximum daily rainfall in the selected region was calculated using a different formula. The rainfall return period is illustrated in Figure 4. The horizontal axis represents the return period in a year while the vertical axis represents the maximum daily rainfall.

**Figure 4**

*Return period in years computed using six different formulas*



From the analysis, the return period of the 40.14 mm event was within the range of 23-44 years as shown in Figure 4. The return period of the 40.14 mm event was 44 years and 23 years according to the plotting point applying Hazen and Weibull method, respectively. The estimate of the return periods from the six methods was in agreement. The results indicated that if the design return period of a hydraulic



infrastructure being designed is less or equal to the data record period, estimation of quintiles by empirical distribution function or plotting point methods is recommendable.

#### 4. Conclusions

This study examined the selection of the best fit probability distribution for various cases of rainfall (average daily rainfall, total rainfall, and maximum daily rainfall) in the Kyrenia region in Northern Cyprus. A 37 different probability distribution models and three goodness-of-fit tests were employed. The results showed that Beta, Dagum, Wakeby, Pareto, Log-Pearson 3, Gen. Extreme Value, and Gen. Gamma (4P) were considered the best model to study the rainfall distributions. Furthermore, six different formulas were used to determine the return periods of maximum daily rainfall.

The results found that the return period of the 40.14 mm event was within the range of 23-44 years. The return period of the 40.14 mm event was 44 years and 23 years according to the plotting point applying Hazen and Weibull method, respectively. Moreover, forecasting rainfall is one of the most important issues in the hydrological cycle.

#### Acknowledgments

The authors wish to thank the Faculty of Engineering especially the Mechanical Engineering Department and the Faculty of Civil and Environmental Engineering especially the Civil Engineering Department for their support during the work validation of this new technology.

#### References

- Alam, M. A., Emura, K., Farnham, C., & Yuan, J. (2018). Best-fit probability distributions and return periods for maximum monthly rainfall in Bangladesh. *Climate*, 6(1), 9. <https://www.mdpi.com/259310>
- FloodList. (2018). Retrieved August 6, 2020, from <http://floodlist.com/europe/cyprus-flash-floods-december-2018>
- Gratien, T., Kundwa, M. J., Bakunzibake, P., Bunani, P., & Habyarimana, J. L. (2019). Development of Rainfall Intensity Duration Frequency (IDF) Curves for Hydraulic Design Aspect. <https://tinyurl.com/yc3x2de4>
- Kassem, Y., & Gokcekus, H. (2020). Water resources and rainfall distribution function: a case study in Lebanon. *Desalination and Water Treatment*, 177, 306-321. [https://www.deswater.com/DWT\\_articles/vol\\_177\\_papers/177\\_2020\\_306.pdf](https://www.deswater.com/DWT_articles/vol_177_papers/177_2020_306.pdf)
- Kratzer, R. (2014). Flooding in Kyrenia - North Cyprus. Retrieved August 6, 2020, from <https://cyprusscene.com/2014/01/18/flooding-in-kyrenia-north-cyprus/>
- LGC News. (2020). North Cyprus News: Flooding Hits Kyrenia - Dam Overflows. Retrieved August 5, 2020, from <https://www.lgcnews.com/flooding-hits-kyrenia-dam-overflows/>
- LGC News. North Cyprus News: Heavy Rain Floods Kyrenia. Retrieved August 5, 2020, from <https://www.lgcnews.com/heavy-rain-floods-kyrenia/>
- Mamoon A., & Rahman, A. (2016). Selection of the best fit probability distribution in rainfall frequency analysis for Qatar. *Natural Hazards*, 86(1), 281-296. <https://link.springer.com/article/10.1007/s11069-016-2687-0>
- Meena R. K., Dubey S. , & Basak S. M. (2019). Use of Different Probability Distribution in Rainfall Analysis for RRS, Pundiari, Cooch Behar. *Int J Curr Microbiol Appl Sci*, 8(10) 1325-1337. <https://doi.org/10.20546/ijcmas.2019.810.155>
- Michaelides, S. C., Tymvios, F. S., & Michaelidou, T. (2009). Spatial and temporal characteristics of the annual rainfall frequency distribution in Cyprus. *Atmospheric Research*, 94(4), 606-615. <https://www.sciencedirect.com/science/article/pii/S016980950900115X>

- Kassem, Y., Gökçekuş, H. & Alijl, N. (2021). Best-fit probability distributions and return periods for daily rainfall in the Kyrenia region, Northern Cyprus. *World Journal of Environmental Research*, 11(2), 39-48. <https://doi.org/10.18844/wjer.v11i2.7189>
- Okoli, K., Mazzoleni, M., Breinl, K., & Di Baldassarre, G. (2019). A systematic comparison of statistical and hydrological methods for design flood estimation. *Hydrology Research*, 50(6), 1665-1678. <https://iwaponline.com/hr/article-abstract/50/6/1665/69899>
- Parchure, A. S., & Gedam, S. K. (2019). Probability distribution analysis of extreme rainfall events in a flood-prone region of Mumbai, India. *Arabian Journal of Geosciences*, 12(11), 1-14. <https://link.springer.com/article/10.1007/s12517-019-4473-2>
- Tfwala, C. M., Van Rensburg, L. D., Schall, R., Mosia, S. M., & Dlamini, P. (2017). Precipitation intensity-duration-frequency curves and their uncertainties for Ghaap plateau. *Climate Risk Management*, 16, 1-9. <https://www.sciencedirect.com/science/article/pii/S2212096317300165>
- UNOOSA. (2010). Flood in Cyprus. Retrieved August 20, 2020, from <http://www.un-spider.org/advisory-support/emergency-support/3288/flood-cyprus>
- Yuan, J., Emura, K., Farnham, C., & Alam, M. A. (2018). Frequency analysis of annual maximum hourly precipitation and determination of best-fit probability distribution for regions in Japan. *Urban climate*, 24, 276-286. <https://www.sciencedirect.com/science/article/pii/S0022169484900088>