DOI: 10.38027/ICCAUA2021243N4

# Statistical Distributions of Precipitation Data in Antalya, Turkey

Asst. Prof. Dr. **Halil Ibrahim Burgan**
*Akdeniz University, Faculty of Engineering, Antalya, Turkey*
*E-mail : burgan@akdeniz.edu.tr*

**Abstract**
The estimation of precipitation data is very important while designing buildings in cities. Because of fast response of rainfall-runoff in urban areas, the stormwater infrastructure systems are developed. Especially in flood prone areas which have flash flood risks, the total precipitation, its return period and its duration should be known. For this aim, statistical distributions of long-term observed precipitation data can be determined. In this study, Antalya district which is near the Mediterranean Sea is selected as study area. The city is very crowded touristic place with more than 2.5 million local people. Antalya is one of the warmest regions in Turkey with an average daily high temperature of 25 degrees centigrade. In 6 months, the average temperatures are over 25 degrees. In the study, the best statistical distributions of precipitation data from the gauging stations near Antalya are investigated. Kolmogorov-Smirnov test is applied to select the best statistical distribution.
***Keywords:*** Antalya; Gauging station; Mediterranean climate; Precipitation; Statistical distribution.

## 1. Introduction

The water that falls to the earth in solid or liquid form from the atmosphere is called precipitation. Liquid precipitation is in the form of rain, while solid precipitation can be in the form of snow, hail, dew, and hoarfrost. The distribution of precipitation is very variable on the earth. There are a lot of factors for this variation. Long term observations are required to be able to define of hydrological procedures as precipitation. As well as natural effects, anthropogenic activities as urbanisation affect negatively the mechanism of precipitation. This phenomenon decreases concentration time in the basin and causes flash floods with erratic rainfall and increasing rainfall intensity. Because of their effects on soil erosion and conservation measures, climatic factors are needed to study and comprehend hydrologic processes on agricultural fields and watersheds. Precipitation is the most variable in time and space of all the meteorological parameters (Apaydin et al., 2006).
Evapotranspiration and infiltration steps start after precipitation event. When they are saturated, the surface runoff starts. The amount water of surface water is easily affected by urbanisation. On the other hand, the estimation of precipitation amount is difficult. Historical records can be used for this aim. There are a lot of studies to take water under control as flood, drought, storage, storage and irrigation. Especially flood and drought extreme events are very important water related disasters. Some precautions are need to be taken.
The distribution of precipitation on a hydrological basin should be defined in terms of water resources planning – management and Hydroelectric Power Plant (HPP) potential. There are some cases for future projection of precipitation regime (Pereira et al., 2017). Because of multivariable precipitation in study regions, the statistical characteristics and the best-fit distributions for precipitation data are investigated. Annual precipitation data from 65 stations in Iran are evaluated to determine suitable probability distribution models for 15, 20, 25 & 30 years return periods (Mahdavi et al., 2010). Precipitation and streamflow data are used to calculate Standardized Precipitation Index (SPI) and Standardized Streamflow Index (SSI), respectively for drought assessments of 121 catchments in United Kingdom. The Tweedie distribution is applied to precipitation data and only for a few sites (Svensson et al., 2017).
The first studies about statistical distributions of precipitation data are widely known and introduced by Mielke (1973). Asymmetric positively-skewed distributions are described the precipitation data. Another popular research is investigated by Juras (1994) for common features definitions of probability distributions for precipitation data. Square-Root-Normal (SRN), Gamma (G) and compound Poisson-Exponential (CPE) distributions are assessed in the study. Markov-chain model is used to understand stochastic processes of maximum daily precipitation (Katz, 1977). The estimation of parameter and statistical tests are made using a large data set for 1-28 February 1930-1969.
The homogeneity tests and trend analyses are applied to precipitation data. The homogeneity test is applied by taking ratios between observed data and some reference values. The reference value is calculated as mean value from surrounding gauged stations. This basic methodology is applied to South-Western Sweden (Alexandersson, 1986). There are a lot of studies on homogeneity tests of precipitation data for country-wide as Southwest of Europe (Gonzalez-Rouco et al., 2001), Turkey (Gokturk et al., 2008), the Netherlands (Buishand et al., 2013), Iraq (Mahmood Agha et al. 2017), Iran (Hosseinzadeh Talaee et al., 2014) and China (Su and Qingxiang, 2014).

Trend analyses show increase, decrease or no change about data for any observation period. There are some studies to determine seasonal or long-term trends on precipitation data characteristics. Karpouzos et al. (2010) is applied trend test methods as Mann-Kendall, Sen's slope to the precipitation data in Greece. A precipitation decrease period for 1987-1993 is detected in the study area. Similarly, trend tests are applied to annual mean and monthly total precipitation data for 1929-1993 observation period in Turkey. Precipitations in January, February and September months and the annual means have trend (Partal and Kahya, 2006).

Besides these studies, statistical analysis and distribution characteristics are investigated in other studies. New techniques as Bayesian approach are tried to fit best distributions of precipitation data. For this aim, mixed Poisson and exponential distributions are evaluated by Korolev et al. (2017). Two study areas have different climatic conditions as Potsdam (Germany) and Elista (Russia) are selected for the application. Multi-parameter probability distributions are widely used to fit precipitation data. Exponential, Beta-P and Weibull distributions are some of the commonly used (Selker and Haith, 1990).

In this study, the best distributions are investigated to fit annual total precipitation data in Antalya City, Turkey. Some widely used statistical distributions as G2: 2-parameter Gamma, GEV: Generalized Extreme Value, GUM: Gumbel, LL2: 2-parameter Log-Logistic, LL3: 3-parameter Log-Logistic, LN2: 2-parameter Log-Normal, LN3: 3-parameter Log-Normal, LP3: 3-parameter Log-Pearson, NOR: Normal, W2: 2-parameter Weibull, W3: 3-parameter Weibull in hydrological sciences are used for the application. Similar statistical distributions are used for low flow analysis (Eris et al., 2019).

## 2. Materials and Methods

Annual total precipitation data is produced from monthly total precipitation data. Then, annual total precipitation is ordered by the first observation period to the last observation period. After the preparation of precipitation data for each gauging station, the statistical characteristics as mean, minimum, maximum, standard deviation, coefficient of variation, coefficient of skewness and autocorrelation coefficient. Then, the data are transferred to statistical software. The best distribution is determined in the statistical analysis. There are many statistical distributions with up to five parameters. In this study, widely used distributions with at most three parameters were preferred. Kolmogorov-Smirnov (KS) test is selected to sort best fitted distributions of data towards worst one. The parameters of selected distributions are calculated. Finally, spatial mapping of best fit distributions is produced covering study area. A basic flow chart is presented in Figure 1.

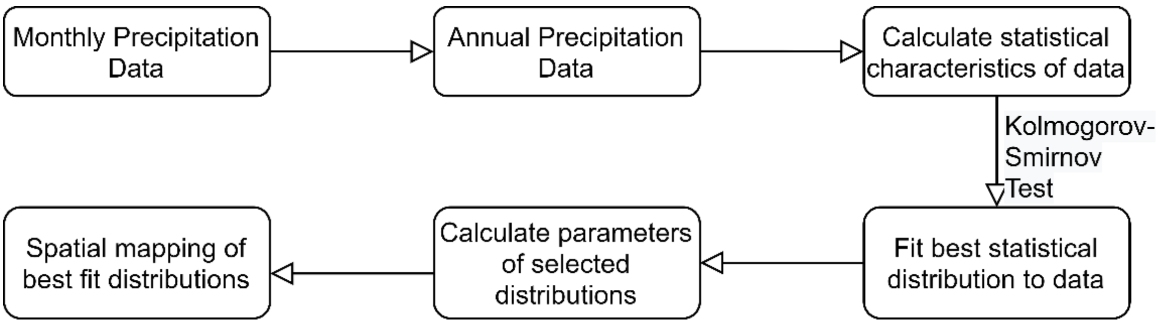


**Figure 1.** A basic flow chart of the methodology.

### 2.1. Kolmogorov-Smirnov Test

The goodness of fit (GOF) tests determine if a random sample is compatible with a theoretical probability distribution function. In other words, these tests demonstrate how well your chosen distribution fits your data. The most known tests are Kolmogorov-Smirnov (KS) test, Anderson-Darling test and Chi-squared test. KS test which is widely used in hydrological sciences is selected. This test determines whether a sample comes from the hypothesized continuos distribution. Its basis is the empirical cumulative distribution function (ECDF). Let's consider a random sample $x_1$, ... , $x_n$ from some distribution with CDF F(x). The empirical CDF is represented by

$$F_n(x) = \frac{1}{n} . [\text{Number of observations} \leq x] \quad (1)$$

The largest vertical difference between the theoretical and empirical cumulative distribution functions is used to calculate the Kolmogorov-Smirnov statistic (D). D can be defined as

$$D = \max_{1\leq i\leq n} \left[F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i)\right] \quad (2)$$

The null and the alternative hypotheses are $H_0$ = the data follow the specified distribution and $H_A$ = the data do not follow the specified distribution. If the test statistic, D, is greater than the critical value obtained from a table, the hypothesis regarding the distributional form is rejected at the given significance level (α). Fixed values of α (0.01, 0.05, and so on) are commonly used to test the null hypothesis ($H_0$) at various significance levels. Most applications use α value of 0.05; however, in some critical industries, a lower value may be used. The standard critical value tables used for this test are only valid when determining whether a data set belongs to a completely specified distribution. When one or more distribution parameters are estimated, the results are conservative: the actual significance level is lower than that provided by the standard tables, and the probability that the fit would be rejected in error is lower. In contrast to fixed α values, the P-value is derived based on the test statistic and signifies the significance level threshold in the sense that the null hypothesis ($H_0$) will be accepted for any values of a smaller than the P-value. For example, if P = 0.025, the null hypothesis will be accepted at all significance levels less than P (i.e. 0.01 and 0.02) but rejected at higher levels, such as 0.05 and 0.1. The P-value is important when the null hypothesis is rejected at all set significance levels and you need to know which level it could be accepted at. For each fitted distribution, P-values based on the Kolmogorov-Smirnov test statistics (D) were determined in this study.

**2.2. Study Area and Data**

There are very different climate conditions in Turkey. Mediterranean climate is effective especially coastal sides of Antalya. The coordinates of Antalya are 36.8874°N latitude and 30.7075°E longitude. The selected study area Antalya, located on Anatolia's southwest coast and bounded by the Taurus Mountains, is the largest Turkish city on the Mediterranean coast outside of the Aegean region, with a population of over 2.5 million people (Figure 2). Antalya which has a high population increment rate by tourism sector and has Mediterranean climate is prone to rapid rainfall – runoff process. Because of this, the precipitation characteristics should be defined truly.

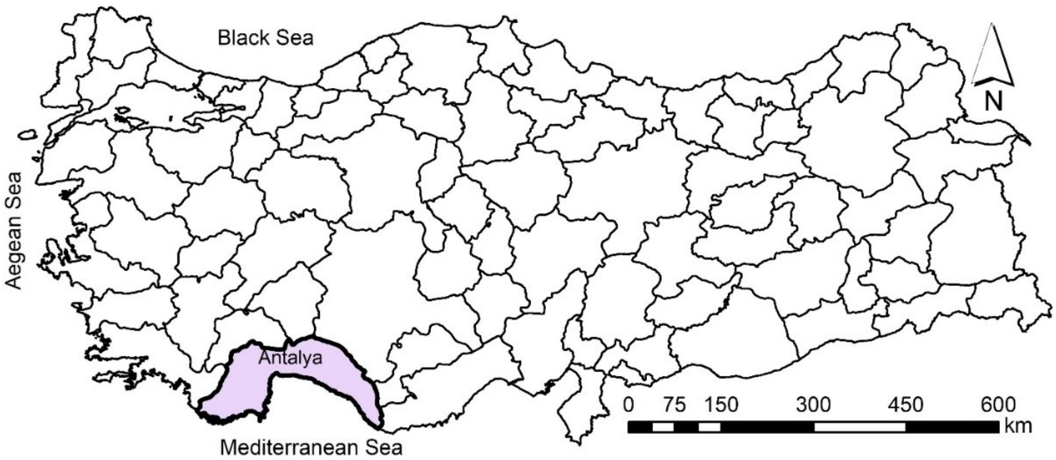


**Figure 2.** Turkish cities and the location of Antalya (study area).

The precipitation observations of ten meteorological stations are used. Eight stations are scattered at the coast side of Antalya, the remaining two stations are in the interior and mountainous side (Figure 3). The coastal sides of Antalya district except 17926 and 17952 gauging stations which are close to Antalya City are assessed in the study. Mediterranean climate is effective in the region. In winter season, there are heavy rainfall. The longest observation period is between 1960 – 2016 (57 years). The shortest observation period is between 2011 – 2016 (6 years) (Table 1).

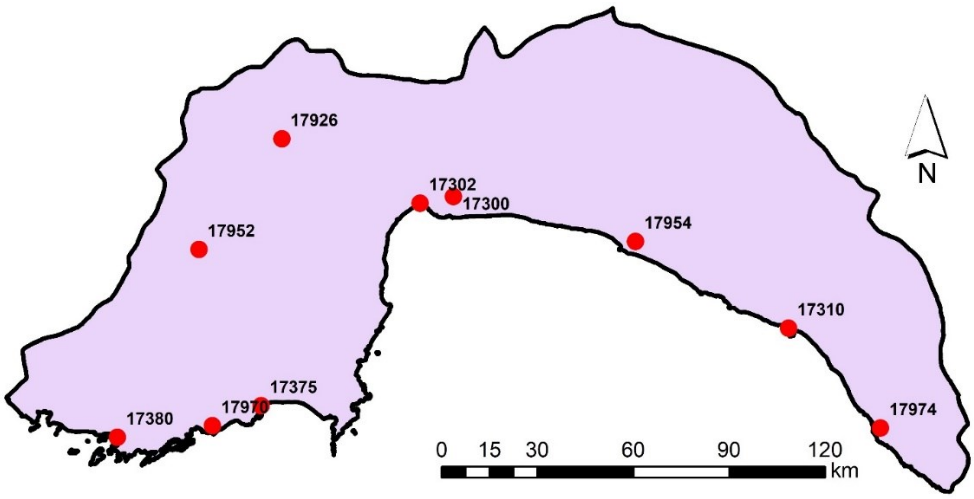


**Figure 3.** The gauging stations used in the study.

**Table 1.** The observation periods and observation years of precipitation data.

| Station ID | Station Name | Observation Period | Observation Years |
|---|---|---|---|
| 17300 | Antalya Airport | 1960 – 2006 | 47 |
| 17302 | Antalya | 2011 – 2016 | 6 |
| 17310 | Alanya | 1960 – 2016 | 57 |
| 17375 | Finike | 1961 – 2016 | 56 |
| 17380 | Kas | 1965 – 1977<br>1982 – 2016 | 42 |
| 17926 | Korkuteli | 1969 – 2011 | 43 |
| 17952 | Elmali | 1960 – 2011 | 52 |
| 17954 | Manavgat | 1963 – 2011 | 49 |
| 17970 | Demre | 1982 – 2011 | 28 |
| 17974 | Gazipasa | 1970 – 2016 | 47 |

The statistical properties of used data are presented in Table 2. The mean value of stations at the coastal side is 961 mm and other two stations have 441 mm mean value. Similarly, standard deviation values for these two interior stations have less than others. The variation coefficients are mostly changed between 0.25-0.30. All stations except antalya station which has the shortest observation period have positive skewness. All autocorrelations are close to zero because of annual precipitation data.

**Table 2.** The statistical properties of precipitation data.

| Station ID | Station Name | Mean | Standard Deviation | Min | Max | $C_v$ | $C_s$ | $\rho_1$ |
|---|---|---|---|---|---|---|---|---|
| 17300 | Antalya Airport | 1119 | 324 | 553 | 1914 | 0.290 | 0.337 | 0.037 |
| 17302 | Antalya | 927 | 233 | 479 | 1253 | 0.251 | -0.959 | 0.195 |
| 17310 | Alanya | 1100 | 300 | 624 | 1786 | 0.272 | 0.447 | -0.067 |
| 17375 | Finike | 955 | 292 | 362 | 1683 | 0.306 | 0.429 | 0.036 |
| 17380 | Kas | 824 | 196 | 429 | 1319 | 0.238 | 0.460 | -0.014 |
| 17926 | Korkuteli | 390 | 116 | 188 | 821 | 0.298 | 1.054 | -0.111 |
| 17952 | Elmali | 491 | 119 | 246 | 777 | 0.242 | 0.090 | 0.066 |
| 17954 | Manavgat | 1112 | 272 | 585 | 1804 | 0.245 | 0.279 | -0.105 |
| 17970 | Demre | 817 | 217 | 410 | 1188 | 0.265 | 0.061 | -0.149 |
| 17974 | Gazipasa | 836 | 222 | 408 | 1431 | 0.265 | 0.738 | -0.121 |

**3. Results**

According to Kolmogorov-Smirnov test, the best fitted distributions are selected. Here, widely used distributions as G2: 2-parameter Gamma, GEV: Generalized Extreme Value, GUM: Gumbel, LL2: 2-parameter Log-Logistic, LL3: 3-parameter Log-Logistic, LN2: 2-parameter Log-Normal, LN3: 3-parameter Log-Normal, LP3: 3-parameter Log-Pearson, NOR: Normal, W2: 2-parameter Weibull, W3: 3-parameter Weibull in hydrological sciences are used for the application. Three best statistical distributions for each station are presented in Table 3. Then, the two or three parameters of the best fitted statistical distributions are calculated in Table 4.

**Table 3.** Three best statistical distributions fitted to data.

| Station ID | Station Name | Distribution 1 | Distribution 2 | Distribution 3 |
|---|---|---|---|---|
| 17300 | Antalya Airport | W2 | GEV | LN3 |
| 17302 | Antalya | GUM | W3 | GEV |
| 17310 | Alanya | W2 | NOR | GEV |
| 17375 | Finike | LL3 | GUM | G2 |
| 17380 | Kas | GUM | LL2 | GEV |
| 17926 | Korkuteli | LL3 | W2 | GEV |
| 17952 | Elmali | GEV | NOR | LP3 |
| 17954 | Manavgat | LL3 | GEV | G2 |
| 17970 | Demre | GEV | LP3 | LN2 |
| 17974 | Gazipasa | GEV | LL3 | GUM |

**Table 4.** The parameters of the best statistical distributions.

| Station ID | Station Name | The Best Distribution | α | β | γ | k | σ | μ |
|---|---|---|---|---|---|---|---|---|
| 17300 | Antalya Airport | W2 | 3.863 | 1218 | | | | |
| 17302 | Antalya | GUM | | | | | 198.6 | 1041 |
| 17310 | Alanya | W2 | 4.249 | 1195 | | | | |
| 17375 | Finike | LL3 | 8.429 | 1378 | -452.6 | | | |
| 17380 | Kas | GUM | | | | | 154.7 | 734.2 |
| 17926 | Korkuteli | LL3 | 9.076 | 552.8 | -173.2 | | | |
| 17952 | Elmali | GEV | | | | -0.271 | 120.0 | 448.1 |
| 17954 | Manavgat | LL3 | 20.90 | 3150 | -2049 | | | |
| 17970 | Demre | GEV | | | | -0.244 | 222.5 | 732.4 |
| 17974 | Gazipasa | GEV | | | | 0.005 | 180.0 | 730.8 |

The best distributions are drawn on the map for each gauging station. According to Figure 4, different statistical distributions are also looking at the coastal side of the study area.

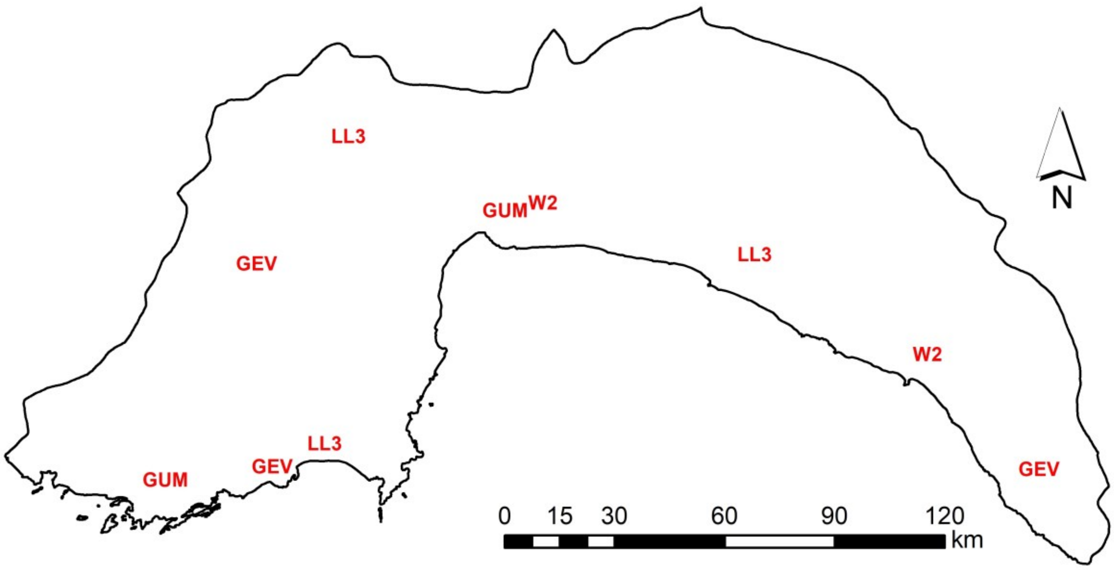


**Figure 4.** Spatial mapping for best fitted distributions.

**4. Discussions**

Main important discussions from the study are:

- The gauging stations have long enough observation periods.
- 17302 (Antalya) which has the shortest observation as years and 17380 (Kas) which is the westernmost station have Gumbel distribution.
- 17300 (Antalya Airport) and 17310 (Alanya) stations which have longest observation periods have 2-parameter Weibull distribution.
- Remaining 6 gauging stations have 3-parameter Loglogistic and Generalized Extreme Value distributions.
- Even so, all gauging stations except 17926 (Korkuteli) which is far away from the coastal side, is located on mountainous region and hasn't Mediterranean climate have Generalized Extreme Value distribution in the best three distributions according to Kolmogorov – Smirnov test.

**5. Conclusions**

Main important conclusions from the study are:

- The observed data can be assessed by dividing observation periods as 10-year to assess urbanization effect on precipitation.
- The tendency of precipitation data can be evaluated to define drought or flood risk management.
- With more gauging stations, homogenous regions can be defined in the basins.
- Rainfall – runoff relationship can be determined for future studies.
- Updated runoff coefficients can be used in terms of design studies considering urbanization.
- Groundwater recharge and surface runoff after extreme rainfall should be investigated with all details in Antalya which karstic geologic formation.

**Acknowledgements**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Conflict of Interests**

The Authors declare no conflict of interest.

**References**

Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology*, 6, 661-675. https://doi.org/10.1002/joc.3370060607

Apaydin, H., Erpul, G., Bayramin, I., Gabriels, D. (2006). Evaluation of indices for characterizing the distribution and concentration of precipitation: A case for the region of Southeastern Anatolia Project, Turkey. *Journal of Hydrology*, 328, 726-732. https://doi.org/10.1016/j.jhydrol.2006.01.019

Buishand, T.A., De Martino, G., Spreeuw, J.N., Brandsma, T. (2013). Homogeneity of precipitation series in the Netherlands and their trends in the past century. *International Journal of Climatology*, 33, 815-833. https://doi.org/10.1002/joc.3471

Eris, E., Aksoy, H., Onoz, B., Cetin, M., Yuce, M.I., Selek, B., Aksu, H., Burgan, H.I., Esit, M., Yildirim, I., Unsal Karakus, E. (2019). Frequency analysis of low flows in intermittent and non-intermittent rivers from hydrological basins in Turkey. *Water Science and Technology: Water Supply*, 19(1), 30–39. https://doi.org/10.2166/ws.2018.051

Gokturk, O.M., Bozkurt, D., Sen, O.L., Karaca, M. (2008). Quality control and homogeneity of Turkish precipitation data. *Hydrological Processes*, 22, 3210-3218. https://doi.org/10.1002/hyp.6915

Gonzalez-Rouco, J.F., Jimenez, J.L., Quesada, V., Valero, F. (2001). Quality control and homogeneity of precipitation data in the Southwest of Europe. *Journal of Climatology*, 14, 964-978. https://doi.org/10.1175/1520-0442(2001)014<0964:QCAHOP>2.0.CO;2

Hosseinzadeh Talaee, P., Kouchakzadeh, M., Some'e, B.S. (2014). Homogeneity analysis of precipitation series in Iran. *Theoretical and Applied Climatology*, 118, 297-305. https://doi.org/10.1007/s00704-013-1074-y

Juras, J. (1994). Some common features of probability distributions for precipitation. *Theoretical and Applied Climatology*, 49, 69-76. https://doi.org/10.1007/BF00868191

Karpouzos, D.K., Kavalieratou, S., Babajimopoulos, C. (2010). Trend analysis of precipitation data in Pieria region (Greece). *European Water*, 30, 31-40.

Katz, R.W. (1977). Precipitation as a chain-dependent process. *Journal of Applied Meteorology*, 16(7), 671-676. https://doi.org/10.1175/1520-0450(1977)016<0671:PAACDP>2.0.CO;2

Korolev, V.Y., Gorshenin, A.K., Gulev, S.K., Belyaev, K.P., Grusho, A.A. (2017). Statistical analysis of precipitation events. *AIP Conference Proceedings*, 1863(1), 090011. https://doi.org/10.1063/1.4992276

Mahdavi, M., Osati, K., Sadeghi, S.A.N., Karimi, B., Mobaraki, J. (2010). Determining suitable probability distribution models for annual precipitation data (A case study of Mazandaran and Golestan Provinces). *Journal of Sustainable Development*, 3(1), 159-168. https://doi.org/10.5539/jsd.v3n1p159

Mahmood Agha, O.M.A., Bagcaci, S.C., Sarlak, N. (2017). Homogeneity analysis of precipitation series in North Iraq. *IOSR Journal of Applied Geology and Geophysics*, 5(3), 57-63. https://doi.org/10.9790/0990-0503025763

Mielke, P.W. (1973). Another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology*, 12(2), 275-280. https://doi.org/10.1175/1520-0450(1973)012<0275:AFODFD>2.0.CO;2

Partal, T., Kahya, E. (2006). Trend analysis in Turkish precipitation data. *Hydrological Processes*, 20(9), 2011-2026. https://doi.org/10.1002/hyp.5993

Pereira, F., Bierhals, E.E., Neris, J.L., Rippel, M., Brazil, C., Salvi, L., Marcal, N. (2017). Evaluation rainfall regime at the hydroelectric power plant toward climate change. *Journal of Contemporary Urban Affairs*, 1(3), 62-65. https://doi.org/10.25034/ijcua.2018.3682

Selker, J.S., Haith, D.A. (1990). Development testing of single-parameter precipitation distributions. *Water Resources Research*, 26(11), 2733-2740. https://doi.org/10.1029/WR026i011p02733

Su, Y., Qingxiang, L. (2014). Improvement in homogeneity analysis method and update of China precipitation data. *Climate Change Research*, 10(4), 276-281. https://doi.org/10.3969/j.issn.1673-1719.2014.04.008