Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 11 NO: 4 PAGE:1026-1041 YEAR: 2022
DOI: 10.17798/bitlisfen.1168077

The Determination of the Most Appropriate Probability Distribution Models for the Meteorological Variables

Musa EŞİT[*1]

[1]A *Civil Engineering Department, Adiyaman University, Adiyaman, Turkey*
(ORCID: 0000-0003-4509-7283)

Keywords: Meteorological Variables, Distribution, Goodness of Fit, Ankara

Abstract
Every component of the hydrological cycle is essential for controlling water supplies and assessing potentially catastrophic events like floods and droughts. The variables of the hydrological system are unexpected and unique to each place. In this paper, the most crucial variables, including precipitation, temperature, relative humidity, and evaporation, are examined for Ankara province. For meteorological parameters, the Lognormal, Log-logistic, Gamma, Weibull, Normal, and Gumbel models are used to find the best suitable distributions. Kolmogorov-Smirnov, Cramers-von Mises, Akaike's Information Criterion, Bayesian Information Criterion, Anderson-Darling, and Maximum Loglikelihood methods are utilized to test these models. The results show that there is a distinct distribution model for each parameter. In particular, it has been determined that the Gumbel distribution is a better model for annual total precipitation, whereas the Normal distribution is a better model for annual minimum temperature. At stations 17130 and 17664, the gamma distribution is observed to be the best fit distribution at annual total precipitation, but station 17128 is found to be the most appropriate for both Log-logistic and normal distribution. Stations 17128, 17130, and 17664 for annual maximum temperature series are fitted with the Normal, Log-logistic, and Lognormal, respectively. Gamma is found to be the best fit when analyzing annual mean temperature for stations 17128 and 17130, whereas Lognormal is selected for station 17664. It is expected that these results will contribute to the planning of water resources projects in the region.

1. Introduction

The hydrological cycle plays a substantial role in the social, economic, and cultural development of any country. The amount and pattern of meteorological variables at a given location are essential factors in a variety of natural and socioeconomic systems, including flood control, water resource management, agriculture, forestry, and tourism [1]. Therefore, it is essential to consider the regime and dynamics of a certain hydrologic phenomenon, particularly the time-based aspects [2]. In addition, the importance of time series analysis is highlighted by the lack of complete understanding of the physical processes involved and the resulting uncertainties in the magnitudes and frequencies of future events [3], [4]. Time series analysis is necessary for developing mathematical models that generate synthetic hydrologic records, detect intrinsic stochastic properties, and forecast hydrologic events of hydrologic variables [5], [6].

In hydrology, it provides the alternative of an acceptable probability distribution function to study rainfall, runoff, and temperature series in various locations [7]. Extreme flooding and rain, however, will cause many people's lives to be disrupted and cost millions of dollars. Hence, the possibility of such an event is necessary for flood control programs, reservoirs, bridges, and other survey management and design staff. The effects of contaminants, unusually low flows, and loads on water must all be taken into consideration in the study of hydrology. As a result,

*Corresponding author: mesit@adiyaman.edu.tr
Received:29.08.2022 Accepted:17.10.2022

they affect the quality and sources of water [8], [9]. Engineering design, planning, and management of water systems and hydraulic structures, including the identification of drought risks, urban planning, and growth forecasting, all largely advantage from understanding the frequency of occurrence of probable values of a random variable through probability distribution. [10], [11].

According to statistical theory, for extremes, the frequency of such occurrences is significantly more influenced by changes in variability (or, more usually, the scale parameter) than by changes in the mean climate (more usually, the location parameter) [12]. The meteorological parameters differ from one country to another as well as from one weather station to another. For example, Khudri [7] discovered that for 50% of the survey stations, generalized gamma four parameter distributions and the generalized extreme value provided the best fit, while no other distribution was consistently detected to be appropriate for the remaining stations in Bangladesh. Unal et al. [13] evaluated the flood flow rates of 2, 5, 10, 25, 50, and 100 years by using the 22-year flow data of 11 AGI stations located in the Gediz Basin. The K-S test was performed to define the most appropriate distribution among Log Normal, Normal, Log Pearson Type III, Gamma, and Gumbel. They determined that the most compatible probability distribution for the flow observation data was Log Pearson Type III. Anli and Anli [14] used the K-S test to find the probability distribution that best fits the 39-year maximum flow data in the Giresun Aksu Basin. As a result of the test, he observed that the distribution most suitable for the annual maximum flow data was the Weibull distribution. Yavuz and Ergül [15] modeled the annual mean flow value of the Eskişehir Porsuk Dam using the Normal, Log Normal, Logistic, Gamma, and Weibull probability distributions. In terms of selecting the best probability distribution for 33 years of data, they discovered that the Weibull probability distribution was more appropriate as a result of the K-S test. Sandalcı [16] employed Normal, Log Normal, and Gumbel probability distributions to determine the flood flow rates for the Akçay Stream, one of the most significant tributaries of the Sakarya River, for recurrence intervals of 5, 10, 25, 50, 100, and 250 years. He utilized the K-S test to identify the probability distribution that was the most consistent and found that the Log Normal distribution was the most entirely compatible. According to Salami [17], it is possible to forecast the amount of rainfall extremely precisely for various durations using a certain probability distribution, even if precipitation varies with time and has unpredictable features. Due to the time and space constraints on rainfall data, accurate estimations are not always possible. Knowledge about extreme rainfall is crucial for many hydrological applications [18]–[20]. Meteorological variables are typically calculated in hydrological research using normal distribution, Pearson type 3, the lognormal distribution, generalized distribution of extreme value (GEV), exponential function, Gumbel distribution, Weibull and Pareto distributions [21], [22]. Haddad [23] investigates various goodness-of-fit (GOF) standards used in various scientific disciplines and examines the benefits and limitations of each GOF in order to compare potential probability density functions (pdfs) to annual maximum temperature data. The annual maximum temperature series is generally best fitted by generalized extreme values and normal distributions. Vivekanandan [24] discovered that the Log-Pearson III distribution was more appropriate for temperature and rainfall data in Hissar, India. On the basis of Australian daily maximum temperatures, Trewin [25] employed a variety of probability distributions; his findings demonstrated that Gaussian distributions represented by various parameters were effective in capturing extremes in the data.

In this study, the most appropriate distribution among the Lognormal, Log-logistic, Gamma, Weibull, Normal, and Gumbel are determined for annual maximum temperature, annual mean temperature, annual minimum temperature, annual total precipitation, annual total evapotranspiration, and annual mean relative humidity for Ankara province. The most suitable distribution is found to perform Cramers-von Mises (CvM), Kolmogorov-Smirnov (KS), Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson-Darling (AD) and Maximum Loglikelihood methods

2. Study Area and Data

Throughout the province's wide territory, there are climatic variations from location to location. The Central Anatolian climate's characteristic steppe climate may be seen in the south, and the temperate and rainy conditions of the Black Sea climate can be observed in the north. In this area, which has a continental climate, winters are cold and summers are hot. The hottest months are July and August, while January is the coldest. The northern and southern parts of the region experience different amounts of precipitation. Ankara displays the climate characteristic of the Central Anatolia Region in the south and the precipitation regime of the Black Sea Region in the north. Fog is a common occurrence and has a negative impact on life because of the region's geography, particularly in the winter. The province

has an average annual temperature of 11.7 °C and 389.1 millimeters of precipitation. The highest recorded temperature was 40.8 °C, while the lowest recorded temperature was -24.9 °C. The highest snow thickness was determined as 30 cm. According to data measured over a long period of time, Ankara's average pressure value is 913.1 mb, the greatest pressure value was 935.0 mb, and the lowest pressure value was 891.0 mb [26], [27]. Time series of meteorological parameters are obtained from the General Directorate of Meteorology Turkey (MGM) for 3 stations located in Ankara (Figure 1). Observations of meteorological station periods and geographic locations are given in Table 1.

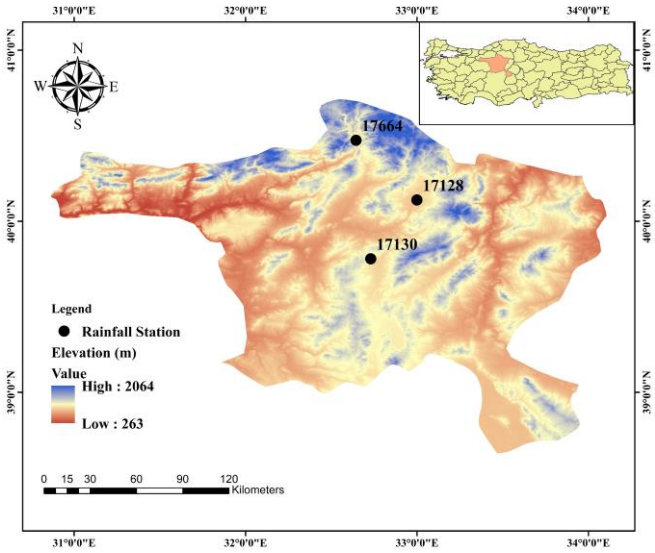

Figure 1. Selected meteorological stations on study area

Table 1. Observation of meteorological stations periods and geographic locations.

Station No	Station Name	Parameters	Longitude	Latitude	Record Years
17128	Ankara Airport	Annual total precipitation (mm) Annual max, min and mean temperature (°C) Annual mean relative humidity (%) Annual total evapotranspiration (mm)	32.999	40.124	1956-2021
17664	Ankara Center	Annual total precipitation (mm) Annual max, min and mean temperature (°C) Annual mean relative humidity (%) Annual total evapotranspiration (mm)	32.644	40.472	1959-2021
17130	Ankara-Kızılcahamam	Annual total precipitation (mm) Annual max, min and mean temperature (°C) Annual mean relative humidity (%) Annual total evapotranspiration (mm)	32.863	39.972	1926-2021

3. Metodology

The decision of probability distribution models is significant for choosing the best-fit probability distribution for a specific area. The chosen distribution models that are frequently employed in assessments of meteorological parameters are given in this section. The approach for parameter estimation is defined, along with numerical and graphical goodness-of-fit assessments for model selection.

3.1. Marginal Probability Distribution

3.1.1 Normal (Gaussian) Distribution

The normal distribution is widely used in the social sciences and plays a significant role in statistics. It illustrates real-valued random variables in the natural sciences when their distribution is not clear [28]. In analyses of annual rainfall and streamflow series, the Gaussian or N distribution is frequently used [29]. The parameters of the normal distribution are mean μ and variance σ^2. For a normal random variable x, the probability density function (pdf), f(x), and cumulative distribution function (cdf), F(x), are given as;

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \quad (1)$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]\right) dx \quad (2)$$

Here, μ represents mean and σ denotes standard deviation. The case where $\mu \to 0$ and $\sigma \to 1$ is referred to the standard normal distribution.

3.1.2.Weibull Distribution

One of the most often used in various fields is the Weibull distribution, which was developed by Swedish physicist Waloddi Weibull. It outlines the quantified failure of a number of different groups of phenomena and components [30]. Pdf (f(x)) and cdf (F(x)) of Weibull distribution are presented as;

$$f(x) = \frac{\gamma}{\alpha}\left(\frac{x-\mu}{a}\right)^{\gamma-1} exp\left(-\left(\frac{x-\mu}{a}\right)^{\gamma}\right) \quad x \geq \mu; \gamma, a > 0 \qquad (3)$$

$$F(x) = 1 - e^{-\left(\frac{x^{\gamma}}{a}\right)} \quad x \geq 0; \gamma > 0 \qquad (4)$$

where α, μ and γ denote scale, location, and shape parameter. If $a = 1$ and $\mu = 0$, it is noted as the standard Weibull distribution, and if $\mu = 0$, it is called the two-parametric Weibull distribution

3.1.3. Gamma Distribution

Due to the gamma distribution's dependence on the normal and exponential distributions, statistics makes extensive use of it. It is described as a two-parametric distribution with continuous probability, just as the logistic distribution. Special cases of the gamma distribution include the exponential, chi-squared, and Erlang distributions [31]. The gamma distribution's fundamental formula for the pdf is written as;

$$f_{(x)} = \frac{\left(\frac{x-\mu}{\beta}\right)^{\gamma-1} exp\left(-\frac{x-\mu}{\beta}\right)}{\beta\Gamma(\gamma)} \quad x \geq \mu; \gamma, \beta > 0 \quad (5)$$

where shape, location, and scale parameter are represented as $\gamma, \beta,$ and μ, respectively. The gamma function Γ, represented as;

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt \qquad (6)$$

The cdf of gamma distribution is;

$$F(x) = \frac{\Gamma_x(\gamma)}{\Gamma(\gamma)} \quad x \geq 0; \gamma > 0 \qquad (7)$$

3.1.4. The Logistic Distribution

The logistic distribution, which has two parameters, is a continuous probability distribution function in statistics. It is typically used in a variety of fields, including logistic regression, logit models, neural networks, finance, sport modeling, physical science, and most recently hydro-climatologic area [32]. Mathematical notation is defined as $X \sim Logistic\ (\mu, s), s > 0$, here μ ($0 \leq \mu \leq \infty$) and s ($s > 0$) represent location parameter and scale parameter respectively. The probability density function (pdf) of logistic distribution is presented by

$$f_X(x) = \frac{e^{-\frac{x-\mu}{s}}}{s\left(1+e^{-\frac{x-\mu}{s}}\right)^2}, \quad -\infty < x < \infty \qquad (8)$$

The cumulative distribution function (cdf) is given as

$$F_X(x) = \int_{-\infty}^{x} x f_X(x) dx = \frac{1}{1+e^{-\frac{x-\mu}{s}}}, \quad -\infty < x < \infty \qquad (9)$$

3.1.5. Lognormal Distribution

For probabilistic design, the lognormal statistical distribution is essential since negative values might occasionally complicate engineering processes. The description of failure rates, fatigue failure, and other circumstances with a wide range of data is generated using lognormal distribution in practice [33]. The lognormal distribution of random variable X, considering expected value μ_x, standard deviation σ_x , given as LN (μ_x, μ_x), is obtained as

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu_Y}{\sigma_Y}\right)^2}, \quad 0 < x < \infty \qquad (10)$$

where, $f_X(x)$ is the pdf of the variable X, and

$$\sigma_y = \sqrt{ln\left(\left(\frac{\sigma_x}{\mu_x}\right)^2 + 1\right)} \qquad (11)$$

and

$$\mu_Y = \ln(\mu_X) - \frac{1}{2}{\sigma_y}^2 \qquad (12)$$

If Y variable indicates a normal distribution, then $X = exp(Y)$. In a same way, if X variable shows a normal distribution, $Y = \ln(X)$.

3.1.6. Gumbel Distribution

The Gumbel model, also known as the extreme value type I distribution, is the most used probabilistic model for addressing extreme events [34] the Gumbel model is presented as

$$F(x) = exp\big(-exp(-(x-u)/\alpha)\big) \qquad (12)$$

where x is the value of the random variable X, u and an are the location and scale parameters, respectively, and F(x) represents for the cumulative distribution function. The pdf of Gumbel distribution is expressed as

$$f(x) = e^{-(x+e^{-x})} \qquad (13)$$

3.2. Goodness-of-Fit Tests

The goodness of fit test (GoF) is used to determine if a variable fits a particular population's distribution. These tests determine whether the distribution is appropriate for random data, to put it another way. The goodness of fit test can be assessed using a variety of techniques. The most often used techniques in this study include Akaike's Information Criterion (AIC) [35], Bayesian Information Criterion (BIC) [36], Cramer-von Mises (CvM), Kolmogorov-Smirnov (K-S) [37], Anderson-Darling (AD) [38], and Maximum Likelihood (ML) methods.

3.2.1. Anderson-Darling Test

Any distribution can be tested using the Anderson-Darlin test [38], which can also be used to determine whether a random variable originated from a population with a particular distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test that gives the tails more weight. The A^2 test statistic for the normal, lognormal, Weibull, and Gumbel distributions can be calculated as

$$A^2 = -n - \left(\frac{1}{n}\right)\sum_{i=1}^{n}(2i-1)[ln(w_i) + ln(1 - w_{n-i+1})] \qquad (14)$$

where n is the sample size and w is the standard normal cdf ($\Phi\left[\frac{(x-\mu)}{\sigma}\right]$)

For Weibull and Gumbel distribution

$$w_i = F(x) = 1 - \exp\left(-\left(\frac{x_i}{n}\right)^{\beta}\right) \qquad (15)$$

where n, β are the model scale and shapes parameters

3.2.2 Kolmogorov-Smirnov (K-S) Test

Kolmogorov-Smirnov test, proposed by Smirnov [39], is weak against variations in distribution tails. Calculations are performed for the directional hypothesis as

$$D^+ = max\{F_{(x)} - G_{(x)}\}$$
$$D^- = min\{F_{(x)} - G_{(x)}\} \qquad (16)$$

where, $F_{(x)}\ and\ G_{(x)}$ indicate the empirical distribution function for the data compared and the combined statistic is given by

$$D = max(|D^+|, |D^-|) \qquad (17)$$

Calculating the asymptotic limiting distribution can indicate the p-value for this theoretical statistic.

$$\lim_{m,n\to\infty} Pr\left\{\sqrt{\frac{mn}{(m+n)}}D_{m,n} \le z\right\} = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}exp(-2i^2z^2)$$

3.2.3. Cramer-von Mises (CvM) Test

The Cramer-von Mises test enables the modeling of a sample vector's $X = (X^1, \dots, X^{nx})$ probability distribution. It examines at whether a random data set and a previously selected candidate probability distribution are compatible [40]. The Cramer-von Mises distance is presented as

$$D = \int_{-\infty}^{\infty}\big(F(x) - F_0(x)\big)^2 dF_0(x) \qquad (19)$$

This test measures the distance between the candidate distribution F and the cumulative distribution function F_0. For testing the hypothesis, $H_o \to F = F_0$. The test statistic mathematically is presented by

$$\widehat{D_N} = \frac{1}{12N} + \sum_{i=1}^{N}\left[\frac{2i-1}{2} - F(x_i)\right]^2 \qquad (20)$$

where the sample size is N and $\widehat{D_N}$ is the asymptotically known probability distribution distance

3.2.4. Information Criterion (AIC) Test

The Akaike's Information Criterion (AIC) [35] is another method for choosing the best model from a group of models. The chosen model displays the smallest difference between the truth and the model. Based on information theory, the results of this test are calculated as

$$AIC = -2\big(ln(likelihood)\big) + 2K \quad (21)$$

Where K stands for the number of free parameters in the model and likelihood is the probability of a variable given a model. AICc is the second-order information criterion taking sample size into account. It is calculated as

$$AICc = -2\big(ln(likelihood)\big) + 2K \times \left(\frac{n}{n-K-1}\right) \quad (22)$$

where n is the sample size.

3.2.5. Bayesian Information Criterion (BIC) Test

The Bayesian Information Criterion (BIC) [36] is another method for choosing appropriate models from a limited number of options. BIC differs from AIC in general, particularly in the second term, which is related to sample size and calculated as

$$BIC = -2logp(L) + plog(n) \quad (23)$$

Where p is the number of predicted parameter and n is the number of the observations. Here, the minimum AIC and BIC are determined as the best distribution.

3.2.6. Maximum likelihood (ML) Method

The maximum likelihood (ML) method identifies an appropriate strategy for parameter prediction problems [41]. The key benefit of utilizing ML is that it extracts all of the valuable data from the input.

Consider a sample $y = [y_1 \ \dots y_i \ \dots \ y_n]$ from the population. Pdf (or the probability density function) of a random variable y_i conditioned on parameters θ is written by $f(y_i, \theta)$. The joint density of n identically and individually disturbed observation is showed as

$$f(y, \theta) = \prod_{i=1}^{N} f(y_i, \theta) = L(\theta|y) \quad (24)$$

And first term $f(y_i, \theta)$ can be given as

$$f(y_i, \theta) = f(y_i, \mu|\sigma^2 = 1) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(y_i-\mu)^2}{2\sigma^2}} \quad (25)$$

It is general practice to study with the Log-Likelihood function.

$$L(\theta|y) = \sum_{i=1}^{N} ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(y_i-\mu)^2}{2\sigma^2}}\right)$$

4. Results and Discussion

4.1. Initial Evaluation and Visualization

To help with data visualization and model selection, Figure 2 displays an initial skewness-kurtosis graph of the unbiased distribution of the meteorological parameters. Due to the limited article page, the data of station 17130 are shown as an example. In this paper, uniform, exponential distributions displayed in the Cullen and Frey graph, developed by Cullen et al. [42] are not selected for best fitting distribution. While the probable beta regions are displayed by larger areas, the probable Lognormal, Gamma, and Weibull regions are depicted by lines. The Cullen and Frey graph shows the kurtosis and squared skewness of the precipitation, temperature, relative humidity, and evapotranspiration series as a blue point representing "observation." According to Figure 2, Given their frequent right-skewed nature and positive skewness, the common right-skewed distributions lognormal, normal, gamma, and Weibull are suggested as potential model distribution options. But, because of the substantial variances in skewness and kurtosis across all distributions, this visualization can only be regarded as suggestive.

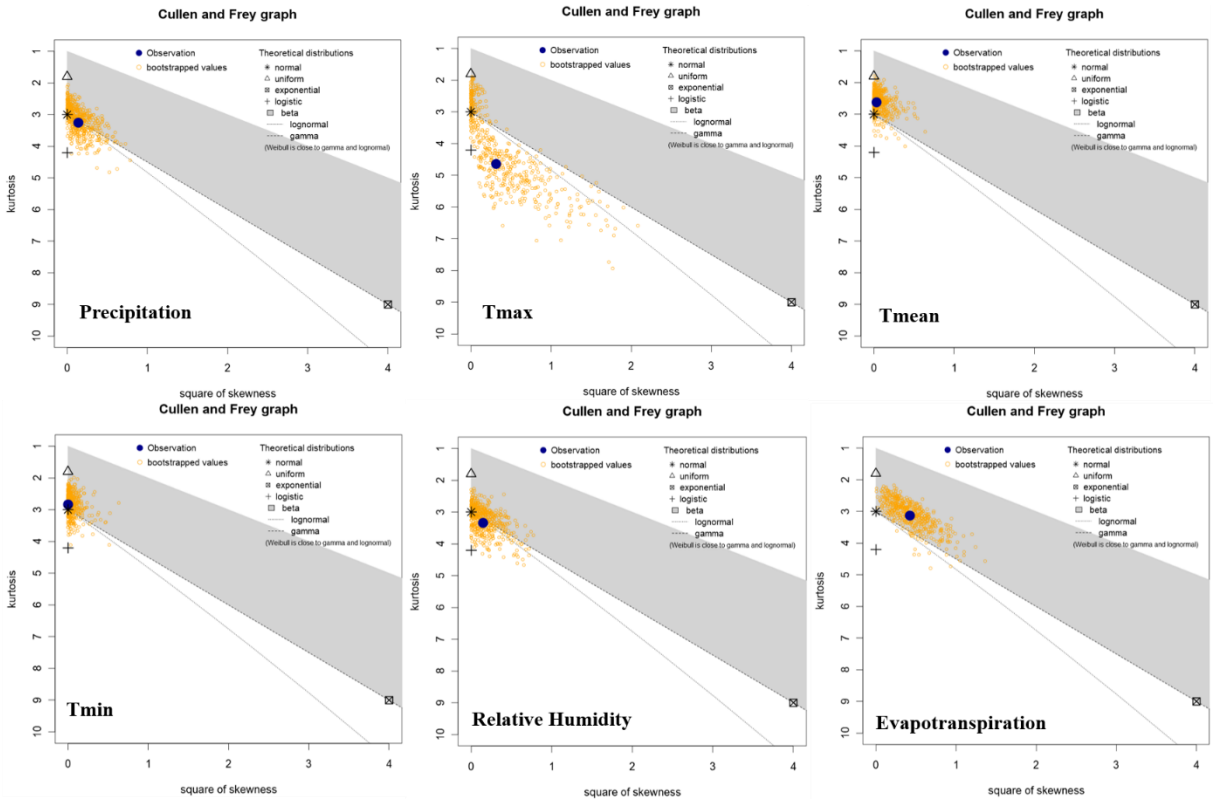

Figure 2. An explanation of precipitation, Tmax, Tmean, Tmin, relative humidity and evapotranspiration series for station 17130 from a normal distribution with estimated bootstrap skewness and kurtosis

Using Assessment-Based Graphs for the GOF

With the help of several graphical functions, the goodness-of-fit of models can be investigated. Figure 3 indicates the theoretical densities of six selected marginal distribution models of precipitation, Tmax, Tmean, Tmin, relative humidity, and evapotranspiration for station 17030. Data analytics requires knowledge of a data's normality or non-normality because it has a significant impact on the algorithms that may be used and how the dataset should be handled. According to Figure 3, log-logistic and Gamma distribution are found the most appropriate models among the selected six distributions for the annual total precipitation series. log-logistic, Gamma and Normal distributions seem to fit the annual maximum, mean, and minimum temperature, respectively. In addition, Log-logistic distribution for annual mean relative humidity and annual total evapotranspiration are observed as more appropriate distribution.

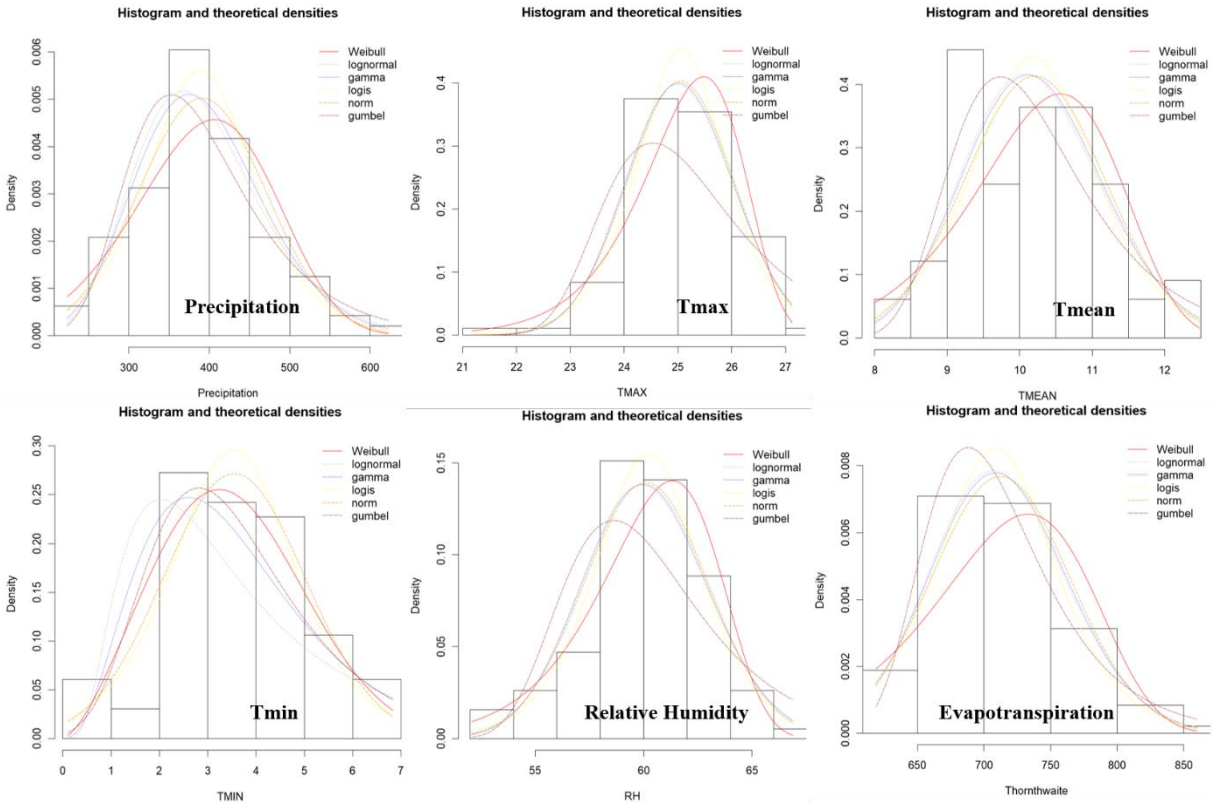

Figure 3. Fitted theoretical densities of six selected marginal distribution models of precipitation, Tmax, Tmean, Tmin, relative humidity and evapotranspiration for station 17030

A probability-probability (P-P) plot is a straightforward graphical technique used to evaluate the accuracy of a forecast prediction and its level of uncertainty. To examine if the dataset series were derived from the six chosen theoretical distributions, quantile-quantile (Q-Q) plots are created for the graphical assessment and visualization of the quality of fit of the selected model distributions. P-P and Q-Q plots are presented for station 17030 in Figures 4 and 5, respectively. The P-P plot compares a uniform distribution to the probability values of the observed meteorological series within the meteorological ensemble, which range from 0 to 1.0. if one of the meteorological series is entirely normally distributed, The P-P plot will be 1:1. The same is also true for the Q-Q plot of the meteorological series. According to Figure 4, P-P plot for all data series is clustered around 1:1 line which means normally distributed. However, the Q-Q plot demonstrates small gaps between empirical quantiles and theoretical quantiles for all marginal selected distributions. Hence, the best fit distribution model is selected considering six different methods and their parameters are predicted by Maximum likelihood method.

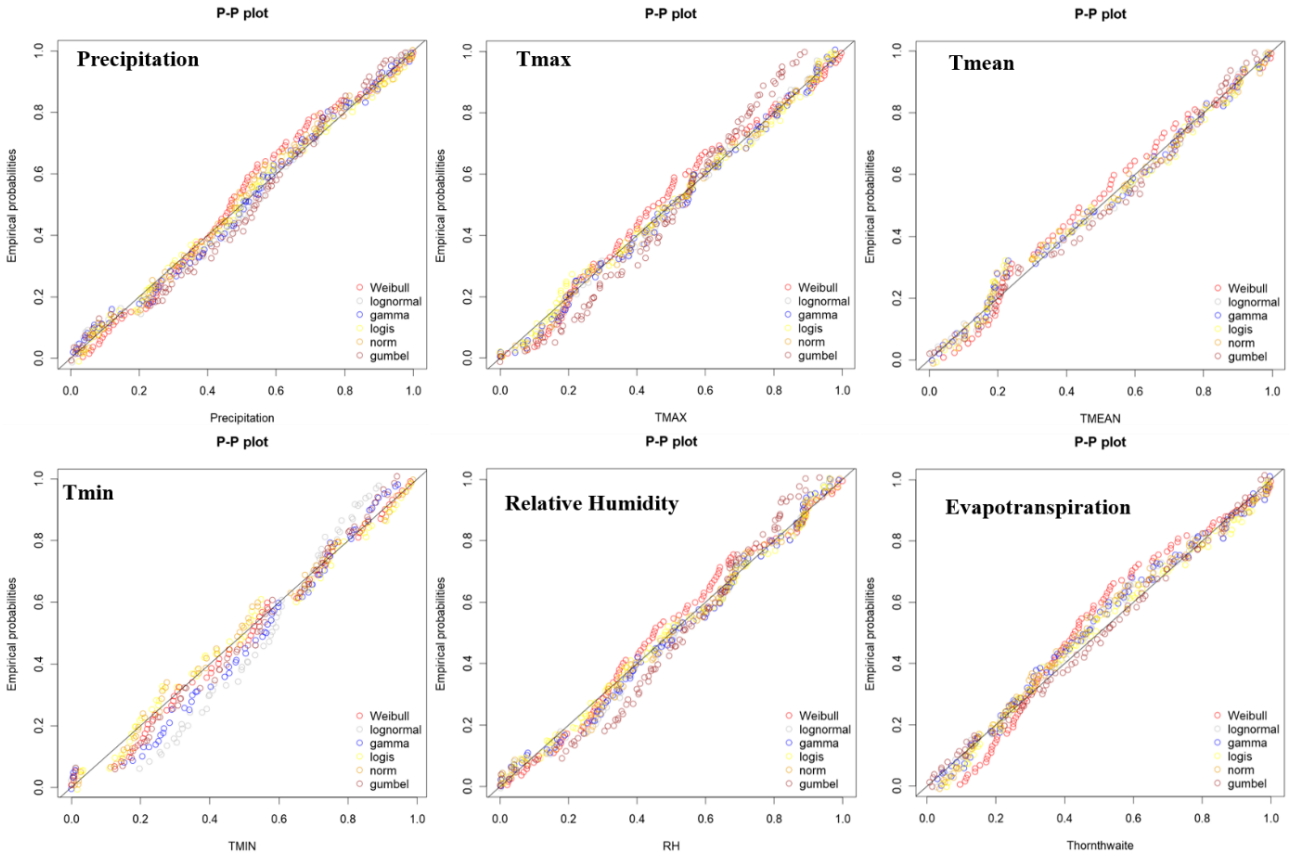

Figure 4. Probability-Probability (P-P) plot for precipitation, Tmax, Tmean, Tmin, relative humidity and evapotranspiration series at station 17030

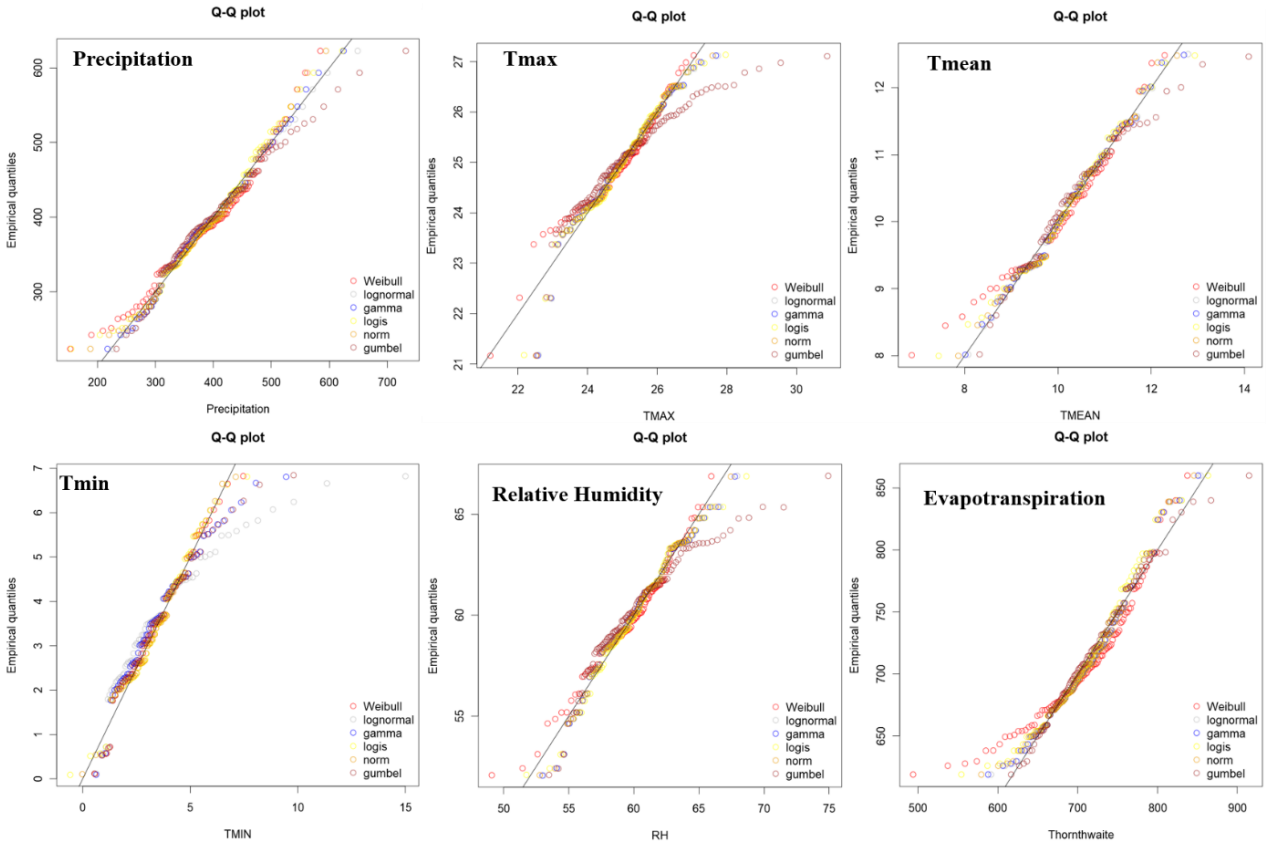

Figure 5. Quantile-Quantile (Q-Q) plot for precipitation, Tmax, Tmean, Tmin, relative humidity and evapotranspiration series at station 17030

Selecting the Best Fitting Distribution Model for Meteorological Parameters

Tables 2, 3, and 4 show the best fitting distribution considering Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling, Akaike's Information Criterion, Bayesian Information Criterion, and Maximum Loglikelihood tests. According to station 17128, the annual total precipitation series fit the logistic and normal distribution. In addition, annual maximum, mean, and minimum temperature series are found to have the best fit distribution as the Normal, Gamma, and Normal models, respectively. While Weibull and Normal distributions are selected as the best fits for annual mean relative humidity, Gumbel distribution is observed as the best fit for annual total evapotranspiration data.

According to the results of station 17130 shown in Table 3, Gamma distribution is selected for annual total precipitation and annual mean temperature. In contrast,the logistic distribution is observed as the best-fit model for maximum yearly temperature, annual mean relative humidity, and annual total evapotranspiration. Furthermore, Normal distribution is fitted to the yearly minimum temperature. At station 17664, the Gamma distribution for annual total precipitation, the Lognormal distribution for maximum yearly and mean temperature, the Normal distribution for annual minimum temperature and annual mean relative humidity, and the Gumbel distribution for annual total evapotranspiration series are selected as the best fit models. Figure 6 shows the best fit model of annual total precipitation for station 17130. It can be inferred from Figure 6 that the Gamma distribution is the most appropriate model for annual total precipitation. Because the Gamma distribution is evenly distributed on the P-P and Q-Q plots, as well as empirical versus theoretical CDFs and theoretical densities.

Table 2. The best fit distribution selection based on GOF tests result for station 17128

Parameters	GOF Tests	Distribution Models					
		Lognormal	logistic	Gamma	Weibull	Normal	Gumbel
Annual total precipitation	K-S	0.1202559	**0.067988**	0.105785	0.095323	0.076522	0.131255
	CvM	0.1285157	**0.035468**	0.090285	0.081791	0.052198	0.214214
	AD	0.788356	**0.235019**	0.551954	0.518097	0.334663	1.34532
	AIC	781.0656	775.7574	778.1951	777.1193	**775.7013**	786.9399
	BIS	785.4449	780.1367	782.5744	781.4986	**780.0807**	791.3192
	ML	-388.5328	-385.879	-387.098	-386.56	**-385.851**	-391.47
Annual max temperature (°C)	K-S	**0.0678967**	0.078759	0.065335	0.085254	0.068022	0.113044
	CvM	0.0595307	0.068418	0.057687	0.069746	**0.054693**	0.128325
	AD	0.3610128	0.447236	0.352026	0.532555	**0.338798**	0.83663
	AIC	181.7881	185.8037	181.6889	185.2186	**181.5621**	189.0137
	BIS	186.1674	190.183	**186.0682**	189.5979	185.9414	193.393
	ML	-88.89405	-90.9019	-88.8445	-90.6093	**-88.7811**	-92.5068
Annual mean temperature (°C)	K-S	0.0857837	0.095759	0.088061	0.077975	0.091818	**0.073505**
	CvM	0.0414913	0.058039	**0.040968**	0.097416	0.044406	0.073541
	AD	**0.2516943**	0.384278	0.254316	0.725902	0.290303	0.503931
	AIC	186.1898	189.4969	**186.1832**	193.4018	186.6707	191.1205
	BIS	190.5691	193.8762	**190.5625**	197.7811	191.05	195.4998
	ML	-91.0949	-92.7484	**-91.0916**	-94.7009	-91.3354	-93.5602
Annual min temperature (°C)	K-S	0.1648745	**0.06674**	0.11403	0.076445	0.067414	0.082981
	CvM	0.5346656	0.044796	0.205697	0.048539	**0.039353**	0.098305
	AD	3.495754	0.313408	1.538355	0.527281	**0.290093**	0.890256
	AIC	283.4662	243.7153	258.4139	246.1259	**242.0193**	250.8365
	BIS	287.8455	248.0946	262.7932	250.5052	**246.3986**	255.2158
	ML	-139.7331	-119.858	-127.207	-121.063	**-119.01**	-123.418
Annual mean relative humidity (%)	K-S	0.1282492	0.097581	0.124824	**0.091136**	0.117984	0.179648
	CvM	0.237001	0.131574	0.22307	**0.129056**	0.197262	0.485943
	AD	1.2798149	0.90904	1.209926	**0.841278**	1.08174	2.584301
	AIC	351.4985	350.723	351.022	352.1487	**350.2268**	364.5233
	BIS	355.8778	355.1024	355.4013	356.528	**354.6061**	368.9026
	ML	-173.7493	-173.362	-173.511	-174.074	**-173.113**	-180.262
Annual total evapotranspiration (mm)	K-S	**0.0824141**	0.096885	0.087604	0.145495	0.097873	0.088929
	CvM	0.1079248	0.111347	0.119375	0.269006	0.144793	**0.040659**
	AD	0.6299967	0.814455	0.690994	1.56304	0.828604	**0.303065**
	AIC	707.9854	713.433	708.5523	720.2482	709.9679	**706.2462**
	BIS	712.3647	717.8123	712.9316	724.6275	714.3472	**710.6255**
	ML	-351.9927	-354.717	-352.276	-358.124	-352.984	**-351.123**

Table 3. The best fit distribution selection based on GOF tests result for station 17130

Parameters	GOF Tests	Distribution Models					
		Lognormal	logistic	Gamma	Weibull	Normal	Gumbel
Annual total precipitation	K-S	0.0573773	0.046804	**0.044613**	0.090082	0.06815	0.079426
	CvM	0.0684747	**0.025833**	0.049171	0.161488	0.063115	0.154016
	AD	0.4135475	**0.23078**	0.297224	0.972485	0.383833	0.910215
	AIC	1115.012	1116.333	**1114.12**	1122.952	1116.305	1119.97
	BIS	1120.14	1121.462	**1119.249**	1128.081	1121.434	1125.098
	ML	-555.5058	-556.167	**-555.06**	-559.476	-556.153	-557.985
Annual max temperature (°C)	K-S	0.0479969	0.056489	0.046959	0.085357	**0.045029**	0.133601
	CvM	0.0346659	0.034154	0.031711	0.138342	**0.027902**	0.480255
	AD	0.3903718	**0.301876**	0.362578	0.912396	0.321002	3.56543
	AIC	277.0407	**272.2233**	276.1155	275.1528	274.4949	314.5435
	BIS	282.1694	**277.352**	281.2442	280.2815	279.6236	319.6722
	ML	-136.5203	**-134.112**	-136.058	-135.576	-135.247	-155.272
	K-S	0.0857837	0.095759	0.088061	0.077975	0.091818	**0.073505**

Annual mean temperature (°C)	CvM	0.0414913	0.058039	**0.040968**	0.097416	0.044406	0.073541
	AD	**0.2516943**	0.384278	0.254316	0.725902	0.290303	0.503931
	AIC	186.1898	189.4969	**186.1832**	193.4018	186.6707	191.1205
	BIS	190.5691	193.8762	**190.5625**	197.7811	191.05	195.4998
	ML	-91.0949	-92.7484	**-91.0916**	-94.7009	-91.3354	-93.5602
Annual min temperature (°C)	K-S	0.1648745	0.06674	0.11403	0.076445	**0.067414**	0.082981
	CvM	0.5346656	0.044796	0.205697	0.048539	**0.039353**	0.098305
	AD	3.495754	0.313408	1.538355	0.527281	**0.290093**	0.890256
	AIC	283.4662	243.7153	258.4139	246.1259	**242.0193**	250.8365
	BIS	287.8455	248.0946	262.7932	250.5052	**246.3986**	255.2158
	ML	-139.7331	-119.858	-127.207	-121.063	**-119.01**	-123.418
Annual mean relative humidity (%)	K-S	0.0757018	**0.050717**	0.073137	0.079285	0.068393	0.133978
	CvM	0.0732437	**0.031862**	0.065636	0.117333	0.053929	0.404769
	AD	0.5548106	**0.282883**	0.500136	0.663733	0.410646	2.686226
	AIC	480.61	**478.3472**	479.8269	480.2198	478.4959	505.1448
	BIS	485.7387	**483.4759**	484.9556	485.3485	483.6246	510.2735
	ML	-238.305	**-237.174**	-237.913	-238.11	-237.248	-250.572
Annual total evapotranspiration (mm)	K-S	0.0442227	0.045106	**0.042755**	0.091797	0.043138	0.109601
	CvM	0.0314271	**0.028243**	0.03046	0.176214	0.029738	0.377656
	AD	0.2833243	**0.246648**	0.272055	1.018951	0.256942	2.710638
	AIC	1029.194	**1027.25**	1028.669	1030.115	1027.74	1061.124
	BIS	1034.323	**1032.379**	1033.798	1035.244	1032.868	1066.253
	ML	-513.5756	-515.89	-514.146	-526.363	-515.503	**-511.536**

Table 4. The best fit distribution selection based on GOF tests result for station 17664

Parameters	**GOF Tests**	**Distribution Models**					
		Lognormal	logistic	Gamma	Weibull	Normal	Gumbel
Annual total precipitation	K-S	0.0888718	0.109425	0.097235	0.104947	0.111849	**0.08782**
	CvM	0.0686	0.10174	0.070188	0.104645	0.087928	**0.068184**
	AD	**0.3503738**	0.582867	0.358499	0.694099	0.487882	0.414419
	AIC	776.3796	780.6651	**776.262**	781.9715	778.0524	778.6765
	BIS	780.6659	784.9514	**780.5483**	786.2577	782.3386	782.9628
	ML	-386.1898	-388.333	**-386.131**	-388.986	-387.026	-387.338
Annual max temperature (°C)	K-S	0.0926571	0.071249	0.095576	0.156059	0.10175	**0.083661**
	CvM	0.0711191	**0.052849**	0.077104	0.30413	0.092019	0.101617
	AD	0.448108	**0.449418**	0.48386	1.776835	0.571741	0.598906
	AIC	**247.0407**	249.3076	247.3349	260.7433	248.1131	250.1337
	BIS	**251.5941**	253.8609	251.8883	265.2966	252.6664	254.687
	ML	**-121.5204**	-122.654	-121.668	-128.372	-122.057	-123.067
Annual mean temperature (°C)	K-S	0.1080861	0.115509	0.112297	0.12538	0.120214	**0.072696**
	CvM	0.0886428	0.111676	0.095552	0.219952	0.112631	**0.083039**
	AD	**0.4885319**	0.637827	0.525097	1.354886	0.624441	0.618401
	AIC	**159.2294**	161.7821	159.4363	169.1497	160.2626	163.4278
	BIS	**163.5156**	166.0684	163.7226	173.436	164.5489	167.714
	ML	**-77.61469**	-78.891	-77.7182	-82.5749	-78.1313	-79.7139
Annual min temperature (°C)	K-S	0.1715909	0.055575	0.127459	0.072962	**0.051689**	0.09267
	CvM	0.5278401	0.024178	0.247025	0.065153	**0.020259**	0.11883
	AD	3.2952866	0.224151	1.654357	0.662941	**0.194787**	0.917298
	AIC	252.1272	222.2973	232.9113	222.4683	**219.6381**	228.4958
	BIS	256.4134	226.5835	237.1976	226.7545	**223.9244**	232.7821
	ML	-124.0636	-109.149	-114.456	-109.234	**-107.819**	-112.248
Annual mean relative humidity (%)	K-S	0.0668988	**0.047915**	0.064259	0.075603	0.059161	0.119563
	CvM	0.0352033	0.033941	0.032352	0.109548	**0.028284**	0.170716
	AD	0.2267838	0.237346	0.21153	0.844457	**0.190637**	1.067333
	AIC	367.808	368.7252	367.7452	378.9523	**367.7756**	377.8099
	BIS	372.3613	373.2785	**372.2985**	383.5057	372.3289	382.3632

	ML	-181.904	-182.363	**-181.873**	-187.476	-181.888	-186.905
Annual total evapotranspiration (mm)	K-S	0.1153349	0.10198	0.120032	0.168421	0.129182	**0.066654**
	CvM	0.1306428	0.124834	0.142467	0.311237	0.167772	**0.048188**
	AD	0.7032216	0.853407	0.763978	1.76613	0.897946	**0.325685**
	AIC	656.7236	660.9225	657.2795	670.3473	658.6173	**655.0107**
	BIS	661.0099	665.2087	661.5658	674.6336	662.9036	**659.297**
	ML	-326.3618	-328.461	-326.64	-333.174	-327.309	**-325.505**

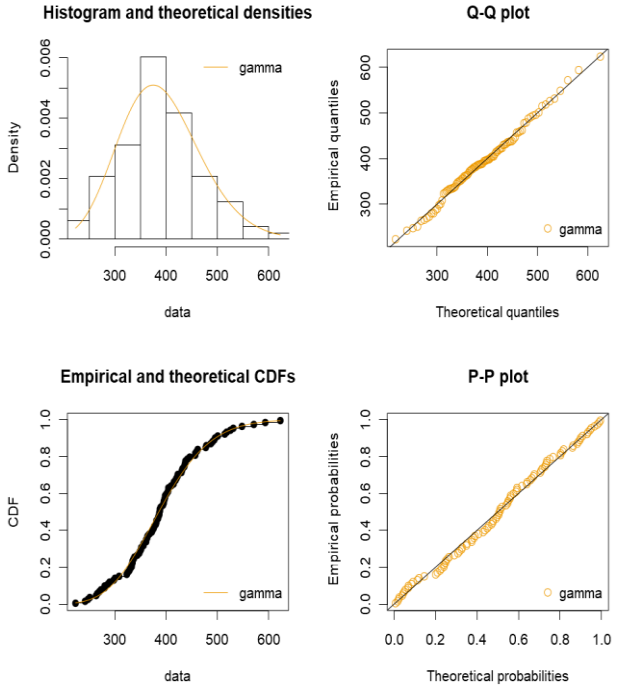

Figure 6. The best fit model for annual total precipitation at station 17130

5. Conclusion

Every component of the hydrological cycle is crucial for managing water supplies and predicting extreme events like floods and droughts. A hydrological system's inputs are unpredictable, thus they are specific to every region. In this study, the most important parameters such as precipitation, temperature, relative humidity, and evaporation are investigated for Ankara province. The most appropriate distributions are determined for meteorological parameters using the Lognormal, Log-logistic, Gamma, Weibull, Normal and Gumbel models. These models are tested by Kolmogorov-Smirnov (KS), Akaike's Information Criterion (AIC), Cramers-von Mises (CvM), Bayesian Information Criterion (BIC), Anderson-Darling (AD), and Maximum Loglikelihood methods.

According to the results, Gamma distribution is found as the best fit distribution for annual total precipitation at stations 17130 and 17664, while Log-logistic and Normal are observed the most appropriate for station 17128. At maximum yearly temperature, all station shows different models. Normal, Log-logistic, and Lognormal are fitted to stations 17128, 17130, and 17664, respectively. When we consider annual mean temperature, for stations 17128 and 17130, Gamma is determined to be the best fit, whereas Lognormal is chosen for station 17664. In addition, Normal distributions are reported as the best fit for all stations at annual minimum temperature series. At the annual mean relative humidity series, Normal is observed to be the best fit for stations 17128 and 17664, while Log-logistic is found as the best fit at station 17130. According to the annual total evapotranspiration series, Gumbel distribution is more appropriate for stations 17128 and 17664, and Log-logistic is fitted better at station 17130. Results indicate that every parameter has a unique distribution model. To be specific, the Gumbel distribution is found more appropriate among other distributions for annual total precipitation, while the Normal distribution is observed more appropriate model at the annual minimum temperature for Ankara province. These findings will serve as a guide for decision-makers on the construction of hydraulic structures for water management in the Ankara province.

Acknowledgments

Acknowledgments are due to state water Works (DSI), general directorate of meteorology (MGM) for providing meteorological data

Contributions of the Authors

Musa Esit: data gathering, hydrometeorological data trend analysis, interpretation of the findings, manuscript writing, and submission

Conflict of Interest Statement

There is no conflict of interest between the authors.

References

[1] U. J. M. A. Alam, K. Emura, C. Farnham, and J. Yuan, "Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh," *Climate*, vol. 6, no. 1, Art. no. 1, Mar. 2018, doi: 10.3390/cli6010009.

[2] M. J. Mamman, O. Y. Martins, J. Ibrahim, and M. I. Shaba, "Evaluation of Best-Fit Probability Distribution Models for the Prediction of Inflows of Kainji Reservoir, Niger State, Nigeria," *Air, Soil and Water Research*, vol. 10, p. 1178622117691034, Jan. 2017, doi: 10.1177/1178622117691034.

[3] I. E. Ahaneku and M. Y. Otache, "Stochastic Characteristics and Modelling of Monthly Rainfall Time Series of Ilorin, Nigeria.," *NONE*, 2014, Accessed: Aug. 27, 2022. [Online]. Available: http://repository.futminna.edu.ng:8080/jspui/handle/123456789/8342

[4] M. Esit, S. Kumar, A. Pandey, D. M. Lawrence, I. Rangwala, and S. Yeager, "Seasonal to multi-year soil moisture drought forecasting," *npj Climate and Atmospheric Science*, vol. 4, no. 1, Art. no. 1, Mar. 2021, doi: 10.1038/s41612-021-00172-z.

[5] E. A. Njoku and D. E. Tenenbaum, "Quantitative assessment of the relationship between land use/land cover (LULC), topographic elevation and land surface temperature (LST) in Ilorin, Nigeria," *Remote Sensing Applications: Society and Environment*, vol. 27, p. 100780, Aug. 2022, doi: 10.1016/j.rsase.2022.100780.

[6] P. Sharma, S. Singh, and S. D. Sharma, "Artificial Neural Network Approach for Hydrologic River Flow Time Series Forecasting," *Agric Res*, Jun. 2021, doi: 10.1007/s40003-021-00585-5.

[7] M. M. Khudri, "Determination of the Best Fit Probability Distribution for Annual Extreme Precipitation in Bangladesh," *European Journal of Scientific Research*, Jan. 2013, Accessed: Aug. 27, 2022. [Online]. Available: https://www.academia.edu/38182722/Determination_of_the_Best_Fit_Probability_Distribution_for_Annual_Extreme_Precipitation_in_Bangladesh

[8] J. Yuan, K. Emura, C. Farnham, and M. A. Alam, "Frequency analysis of annual maximum hourly precipitation and determination of best fit probability distribution for regions in Japan," *Urban Climate*, vol. 24, pp. 276–286, Jun. 2018, doi: 10.1016/j.uclim.2017.07.008.

[9] E. Eris *et al.*, "Frequency analysis of low flows in intermittent and non-intermittent rivers from hydrological basins in Turkey," *Water Supply*, vol. 19, no. 1, pp. 30–39, Feb. 2019, doi: 10.2166/ws.2018.051.

[10] J. Liu, C. D. Doan, S.-Y. Liong, R. Sanders, A. T. Dao, and T. Fewtrell, "Regional frequency analysis of extreme rainfall events in Jakarta," *Nat Hazards*, vol. 75, no. 2, pp. 1075–1104, Jan. 2015, doi: 10.1007/s11069-014-1363-5.

[11] M. I. Yuce and M. Esit, "Drought monitoring in Ceyhan Basin, Turkey," *Journal of Applied Water Engineering and Research*, vol. 0, no. 0, pp. 1–22, Jun. 2021, doi: 10.1080/23249676.2021.1932616.

[12] R. W. Katz and B. G. Brown, "Extreme events in a changing climate: Variability is more important than averages," *Climatic Change*, vol. 21, no. 3, pp. 289–302, Jul. 1992, doi: 10.1007/BF00139728.

[13] H. B. Unal, S. Asik, M. Avci, S. Yasar, and E. Akkuzu, "Performance of water delivery system at tertiary canal level: a case study of the Menemen Left Bank Irrigation System, Gediz Basin, Turkey," *Agricultural Water Management*, vol. 65, no. 3, pp. 155–171, Mar. 2004, doi: 10.1016/j.agwat.2003.10.002.

[14] A. S. Anli and A. S. Anli, "Giresun Aksu Havzası Maksimum Akımlarının Frekans Analizi," *Akdeniz Üniversitesi Ziraat Fakültesi Dergisi*, vol. 19, no. 1, Art. no. 1, Mar. 2006.

[15] H. Yavuz and S. Erdoğan, "Spatial Analysis of Monthly and Annual Precipitation Trends in Turkey," *Water Resour Manage*, vol. 26, no. 3, pp. 609–621, Feb. 2012, doi: 10.1007/s11269-011-9935-6.

[16] M. Sandalci, "Flood Frequency Analysis of Akçay Stream," *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 22, no. 5, Art. no. 5, 2018, doi: 10.16984/saufenbilder.402190.

[17] A. W. Salami, "Best-fit Probability Distribution model for peak daily rainfall of selected Cities in Nigeria," *New York Science Journal*, Jan. 2009, Accessed: Aug. 27, 2022. [Online]. Available: https://www.academia.edu/1593242/Best_fit_Probability_Distribution_model_for_peak_daily_rainfall_of_selected_Cities_in_Nigeria

[18] M. T. Amin, M. Rizwan, and A. A. Alazba, "A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan," *Open Life Sciences*, vol. 11, no. 1, pp. 432–440, Jan. 2016, doi: 10.1515/biol-2016-0057.
[19] H. Sun, G. Wang, X. Li, J. Chen, B. Su, and T. Jiang, "Regional frequency analysis of observed sub-daily rainfall maxima over eastern China," *Adv. Atmos. Sci.*, vol. 34, no. 2, pp. 209–225, Feb. 2017, doi: 10.1007/s00376-016-6086-y.
[20] G. Chen, J. Norris, J. D. Neelin, J. Lu, L. R. Leung, and K. Sakaguchi, "Thermodynamic and Dynamic Mechanisms for Hydrological Cycle Intensification over the Full Probability Distribution of Precipitation Events," *Journal of the Atmospheric Sciences*, vol. 76, no. 2, pp. 497–516, Feb. 2019, doi: 10.1175/JAS-D-18-0067.1.
[21] N. Boudrissa, H. Cheraitia, and L. Halimi, "Modelling maximum daily yearly rainfall in northern Algeria using generalized extreme value distributions from 1936 to 2009," *Meteorological Applications*, vol. 24, no. 1, pp. 114–119, 2017, doi: 10.1002/met.1610.
[22] M. Douka, T. S. Karacostas, E. Katragkou, and C. Anagnostolpoulou, "Annual and Seasonal Extreme Precipitation Probability Distributions at Thessaloniki Based Upon Hourly Values," in *Perspectives on Atmospheric Sciences*, Cham, 2017, pp. 521–527. doi: 10.1007/978-3-319-35095-0_75.

[23] K. Haddad, "Selection of the best fit probability distributions for temperature data and the use of L-moment ratio diagram method: a case study for NSW in Australia," *Theor Appl Climatol*, vol. 143, no. 3, pp. 1261–1284, Feb. 2021, doi: 10.1007/s00704-020-03455-2.
[24] N. Vivekanandan, "Comparison of probability distributions in extreme value analysis of rainfall and temperature data," *Environ Earth Sci*, vol. 77, no. 5, p. 201, Mar. 2018, doi: 10.1007/s12665-018-7356-z.
[25] B. C. Trewin, "Extreme temperature events in Australia," 2001. Accessed: Aug. 27, 2022. [Online]. Available: https://scholar.google.com/scholar_lookup?title=Extreme+temperature+events+in+Australia&author=Trewin%2C+Blair+C.&publication_year=2001
[26] S. Sensoy and M. Demircan, "Climate of Turkey," Mar. 2016.
[27] A. Danandeh Mehr, A. U. Sorman, E. Kahya, and M. Hesami Afshar, "Climate change impacts on meteorological drought using SPI and SPEI: case study of Ankara, Turkey," *Hydrological Sciences Journal*, vol. 65, no. 2, pp. 254–268, Jan. 2020, doi: 10.1080/02626667.2019.1691218.
[28] A. Lyon, "Why are Normal Distributions Normal?," *Br J Philos Sci*, vol. 65, no. 3, pp. 621–649, Sep. 2014, doi: 10.1093/bjps/axs046.
[29] R. D. Markovic, "Probability functions of best fit to distributions of annual precipitation and runoff," 1965. Accessed: Aug. 27, 2022. [Online]. Available: https://scholar.google.com/scholar_lookup?title=Probability+functions+of+best+fit+to+distributions+of+annual+precipitation+and+runoff&author=Markovic%2C+Radmilo+D.&publication_year=1965
[30] C.-D. Lai, D. N. Murthy, and M. Xie, "Weibull Distributions and Their Applications," in *Springer Handbook of Engineering Statistics*, H. Pham, Ed. London: Springer, 2006, pp. 63–78. doi: 10.1007/978-1-84628-288-1_3.
[31] K. J, O. Ngesa, and G. Orwa, "On Generalized Gamma Distribution and Its Application to Survival Data," *International Journal of Statistics and Probability*, vol. 8, no. 5, pp. 85–102, 2019.
[32] R. Kissell and J. Poserina, *Optimal Sports Math, Statistics, and Fantasy*. Academic Press, 2017.
[33] K.-H. Chang, "Chapter 3 – Solid Modeling," 2015. doi: 10.1016/B978-0-12-382038-9.00003-X.
[34] E. Castillo, *Extreme Value Theory in Engineering*. Elsevier, 2012.
[35] H. Akaike, "'An information criterion (AIC).,'" *Math Sci*, vol. 14 (153), pp. 5–7, 1976.
[36] M. Stone, "Comments on Model Selection Criteria of Akaike and Schwarz," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 41, no. 2, pp. 276–278, 1979.
[37] N. Smirnov, "Table for Estimating the Goodness of Fit of Empirical Distributions," *The Annals of Mathematical Statistics*, vol. 19, no. 2, pp. 279–281, 1948.

[38] M. A. Stephens, "EDF Statistics for Goodness of Fit and Some Comparisons," *null*, vol. 69, no. 347, pp. 730–737, Sep. 1974, doi: 10.1080/01621459.1974.10480196.
[39] N. Smirnov, "Estimate of deviation between empirical distribution functions in two independent samples," *Bulletin Moscow University*, vol. 2 (2), no. 3–16, 1939.
[40] F. Laio, "Cramer–von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters," *Water Resources Research*, vol. 40, no. 9, 2004, doi: 10.1029/2004WR003204.
[41] C. Schwarz, "Sampling, Regression, Experimental Design and Analysis for Environmental Scientists, Biologists, and Resource Managers," Mar. 2011.
[42] A. C. Cullen, H. C. Frey, and C. H. Frey, *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Springer Science & Business, 1999.