

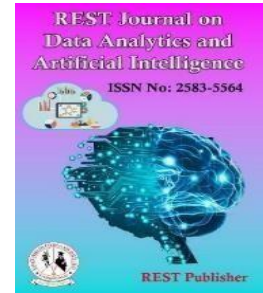
REST Journal on Data Analytics and Artificial Intelligence

Vol: 3(4), December 2024

REST Publisher; ISSN: 2583-5564

Website: <http://restpublisher.com/journals/jdaai/>

DOI: <https://doi.org/10.46632/jdaai/3/4/8>



Modelling Annual Maximum Daily Rainfall in Kolhapur District: A Comparison of Gumbel, Weibull, and Generalized Extreme Value Distributions with Return Period Estimates

Umesh Rajendra Yeole*

Pratap College, Amalner, Maharashtra, India.

*Corresponding author Email: uryeole@gmail.com

Abstract: This study models the annual maximum daily rainfall (AMDR) of Kolhapur district by the probability distributions like Gumbel, Weibull, and Generalized Extreme Value (GEV). The study uses 101 years of datasets collected over 17 stations. The maximum likelihood technique of estimation has been used for estimating the parameters of the model, whereas the Kolmogorov-Smirnov test is used to assess the goodness of fit of the fitted model. To identify the best-fitted distribution, the mean square error (MSE) criterion is used. It is found that the generalized extreme value distribution (GEV) has been shown to have a minimum value of MSE. The highest AMDR value of 780 mm is also captured by the GEV distribution, which was impossible with conventional distributions. For each of the aforementioned probability distributions, the return period and return values with a 95% confidence interval are also calculated. The Statistical Analysis is performed using R software (version 4.5.0).

Keywords: Annual maximum daily rainfall, Gumbel Distribution, Generalized Extreme Value distributions, Return period, Return Value, Weibull distribution, etc.

1. INTRODUCTION

The Kolhapur district, located between Latitude 15°73'N to 17°11'N and Longitude 73°75'E to 74°70'E, surrounded by the Sahyadri, is known for its uneven rainfall patterns. The major rivers, namely Warana, Krishna, and Panchganga, running through the district, become flood-prone during heavy rainfall. In the past, the district has received an extreme rainfall on several instances. Notable extreme rainfall events in the region include: 1652 mm recorded at Gaganbawda station between 24th July and 2nd August 2005; over 1500 mm of continuous rainfall in a week between 5th and 11th August 2019; 481.5 mm at Kasaba Walawe, 433.3 mm at Aavali Bk, and 425.5 mm at Bindri station on 23rd July 2021; and 245 mm at Patgaon station on 16th July 2024. Estimating the annual maximum daily rainfall (AMDR) value that a particular region may receive for the next 100 years (say) may help the administration to take appropriate actions (Coles et al., 2003). Thus, the study of these extreme events is an important tool for estimating the expected rainfall amount (Subudhi, 2007). Several researchers have used the exponential, Weibull, Lognormal, Gumbel, Log-Pearson (LP), and Generalized Extreme Value (GEV) distributions to model the AMDR data. In the Global context, Kouakou and Saralees modeled the AMDR of twelve stations located across Senegal by the GEV distribution (Ndri & Nadarajah, 2022). In India, a study conducted over the Ludhiana region of Punjab used the log-normal, Gumbel, and Log-Pearson type-III distributions to model the AMDR data for 38 years (1970 to 2007) and found that the Log-Pearson type III distribution is the best-fitted distribution (Kumar & Bhardwaj, 2015), another study for the Jhalrapatan area of Rajasthan, has used Weibull, Pearson type -III, and Gumbel distributions to model the AMDR over 41 years (1973 to 2013) and found that the Log-Pearson type-III distribution is the best-fitted distribution (Bhim Singh & Singh, 2012), a recent study conducted for the Afzalpur and Kalaburagi meteorological station located in Karnataka, used the six probability distributions viz. exponential, Extreme value type -I, Extreme value type -II, GEV, Generalized Pareto, and 2-parameter lognormal distributions to model the AMDR data of 49 years (1970 to 2018) and found that GEV as the best fitted distribution (Vivekanandan, 2012). A study conducted across the

four stations, viz. Bhira, Khalapur, Mangaon, and Murud stations located in Raigad district of Maharashtra have used the Gumbel (EV-I) and Frechet(EV-II) distributions to model the AMDR data of 29 years (1968 to 1996) and found the EV-I distribution as the best fit(Vivekanandan, 2021).Although several studies have modeled Annual Maximum Daily Rainfall (AMDR) across Maharashtra, Kolhapur District, a region highly vulnerable to extreme rainfall and recurrent flooding, remains critically understudied. This research fills this gap by analyzing Kolhapur's 101-year AMDR record (1913–2013) using advanced probability distributions, including the Gumbel, Generalized Extreme Value (GEV), and Weibull models. By evaluating these distributions through rigorous statistical tests and return period analysis, this study provides the first comprehensive probabilistic assessment of extreme rainfall risks specific to Kolhapur, offering vital insights for flood mitigation and infrastructure planning. The results obtained in the study would help to answer questions like “What is the maximum amount of daily maximum rainfall Kolhapur district could receive?”, What will be the expected value of AMDR in the next say 25, 50, 100 years, etc, for Kolhapur district. This paper has been organized as follows. Section 2 describes the data used in the analysis. Section 3 describes the Statistical methods and terminologies used. Section 4 discusses the fitting of probability distributions, Weibull, Gumbel, and GEV to the rainfall data, goodness of fit, and the estimation of the return period. In section 5, conclusions obtained from the analysis are discussed.

2. MATERIALS AND METHODS

Data Description

The gridded dataset used for the analysis downloaded from the Indian Meteorological Department (IMD website). The raw dataset provides the daily rainfall (in mm) for 101 years (from 1913 to 2013) over 17 stations, viz. Ajra, Bhudargad, Chandgad, Gadhingalaj, Gaganbawada, Gargoti, Hatkanangale, Kagal, Kapashi, Karveer, Kolhapur, Kurundwad, Panhala, Radhanagari, Shahuwadi, Shirol and Wadgaon. The study variable AMDR is derived in the following two steps

Step 1: Calculation of the maximum daily rainfall (MDR) over a district

In this step, the maximum value of rainfall on a particular day across all 17 stations is calculated. This is referred to as the maximum daily rainfall. For instance, if on August 2, 2005, the maximum rainfall value among the 17 stations is 150mm, it is regarded as the daily maximum rainfall (MDR) for the Kolhapur district on that day. As a result, there will be either 365 or 366 values of MDR in a year

Step 2: Calculation of Annual Maximum Daily Rainfall (AMDR)

The maximum daily rainfall (MDR) calculated in step 1 serves as the basis for determining the Annual Maximum Daily Rainfall (AMDR). The highest of the maximum daily rainfall (MDR) observed within a year is specifically chosen as the annual maximum daily rainfall (AMDR) for that particular year, representing the peak daily rainfall recorded. Each year is associated with a single AMDR value, leading to a total of 101 distinct AMDR values over the span of 101 years.

Methods

Extreme value distributions (EVD) are used widely for the modelling of extreme values of variables. In this paper, Gumbel, Weibull, and Generalized extreme value distributions have been fitted to the data. The method of maximum likelihood estimation is considered as good method for the modeling of meteorological data using the Weibull distribution by Ikbali et.al.(Ikbali et al., 2022). For fitting probability models, R-software (version 4.5.0) is used (Team, 2025)

Let X denote a continuous random variable that represents Annual Maximum Daily Rainfall (AMDR).

Gumbel Distribution

The probability density function (p.d.f) X with location parameter $\theta \in \mathbb{R}$ and scale parameter $\alpha > 0$ is

$$f(x; \theta, \alpha) = \frac{1}{\alpha} e^{-(z+e^{-z})} \quad ; \quad x \in \mathbb{R} \text{ where } z = \frac{x - \theta}{\alpha}$$

The distribution function is

$$F(x; \theta, \alpha) = e^{(-e^{-z})}$$

The p^{th} quantile is

$$x_p = \theta - \alpha \log[-\log(p)]$$

Let $X_1, X_2, \dots, X_n$ denote a random sample from a distribution having p.d.f $f(x; \theta, \alpha,)$ and $x_1, x_2, \dots, x_n$ denotes its respective observed values. The log-likelihood function is

$$-\sum_{i=1}^n \frac{x_i - \theta}{\alpha} - \sum_{i=1}^n e^{-\left(\frac{x_i - \theta}{\alpha}\right)} - n \ln(\alpha)$$

The maximum likelihood estimate of α is the solution of the following equation

$$\bar{x} - \frac{\sum_{i=1}^n x_i \exp\left(-\frac{x_i}{\alpha}\right)}{\sum_{i=1}^n \exp\left(-\frac{x_i}{\alpha}\right)} - \alpha = 0 \quad (1)$$

The maximum likelihood estimate of θ is

$$\hat{\theta} = -\hat{\alpha} \ln \left[\frac{n}{\sum_{i=1}^n \exp\left(-\frac{x_i}{\hat{\alpha}}\right)} \right]$$

The equation 1 is solved by using the *uniroot* function in R software, which provides the maximum likelihood estimate (mle) of α . This mle of the parameter α is used to obtain the maximum likelihood estimate (mle) of the parameter θ (Singh, 1998).

Weibull Distribution

The probability density function (p.d.f) of a two-parameter Weibull distribution with scale parameter $\alpha > 0$ and shape parameter $\beta > 0$ is

$$f(x; \alpha, \beta) = \left(\frac{\beta}{\alpha}\right) \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right], \quad x > 0$$

The distribution function is

$$F(x; \alpha, \beta) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right]$$

The p^{th} quantile is

$$x = \alpha [-\ln(1 - p)]^{\frac{1}{\beta}}$$

Let $X_1, X_2, \dots, X_n$ be a random sample from a distribution with p.d.f $f(x; \alpha, \beta)$ and $x_1, x_2, \dots, x_n$ denote its respective observed values. The log-likelihood function is

$$\ln L(\alpha, \beta) = n \ln \beta - n \beta \ln \alpha - \frac{1}{\alpha^\beta} \sum_{i=1}^n x_i^\beta + (\beta - 1) \sum_{i=1}^n \ln x_i$$

The maximum likelihood estimate of β is the solution of the likelihood equation

$$\frac{1}{\beta} - \frac{\sum_{i=1}^n x_i^\beta \ln x_i}{\sum_{i=1}^n x_i^\beta} + \frac{1}{n} \sum_{i=1}^n \ln x_i = 0 \quad 2$$

The maximum likelihood estimate of α is

$$\hat{\alpha} = \left(\frac{1}{n} \sum_{i=1}^n x_i^{\hat{\beta}} \right)^{\frac{1}{\hat{\beta}}}$$

The solution to equation 2 is obtained using the *uniroot* function in R, which provides the maximum likelihood estimate (mle) of the parameter β . The maximum likelihood estimate (mle) of the parameter β is used to obtain the maximum likelihood estimate of α (Ikbal et al., 2022)

Generalized Extreme Value Distribution

The probability density function (p.d.f) of the Generalized Extreme Value Distribution having location parameter $\theta \in R$, scale parameter $\alpha > 0$, and shape parameter $\beta \in R$ is

$$f(x; \theta, \alpha, \beta) = \alpha^{-1} [t(x)]^{\beta+1} e^{-t(x)} \quad x \in \begin{cases} \left(\theta - \frac{\alpha}{\beta}, \infty\right), & \beta > 0 \\ (-\infty, \infty), & \beta = 0 \\ \left(-\infty, \theta - \frac{\alpha}{\beta}\right), & \beta < 0 \end{cases}$$

$$\text{where } t(x) = 1 + \beta \left(\frac{x - \theta}{\alpha}\right)^{-\left(\frac{1}{\beta}\right)} \quad \text{if } \beta \neq 0$$

$$= e^{-\left(\frac{x - \theta}{\alpha}\right)} \quad \text{if } \beta = 0$$

The distribution function is

$$F(x; \theta, \alpha, \beta) = \begin{cases} \exp\left\{-\left[1 - \beta \left(\frac{x - \theta}{\alpha}\right)^{\frac{1}{\beta}}\right]\right\} & \beta \neq 0 \\ \exp\left\{-\exp\left(-\frac{x - \theta}{\alpha}\right)\right\} & \beta = 0 \end{cases}$$

The p^{th} quantile is $x_p = \theta + \frac{\alpha\{1 - [-\log(p)]^{\beta}\}}{\beta}$

Let $X_1, X_2, \dots, X_n$ denote a random sample from a distribution having p.d.f $f(x; \theta, \alpha, \beta)$ and $x_1, x_2, \dots, x_n$ denote its respective observed values. The Log-likelihood function is

$$\ln L(x; \theta, \alpha, \beta) = -n \ln \alpha - (1 - \beta) \sum_{i=1}^n y_i - \sum_{i=1}^n e^{-y_i}$$

$$\text{where } y_i = -\beta^{-1} \ln \left\{ -\beta \left(\frac{x_i - \theta}{\alpha}\right) \right\}$$

The maximum likelihood estimates of θ, α , and β can be obtained by solving the following equations.

$$\frac{\partial \ln L(x; \theta, \alpha, \beta)}{\partial \theta} = 0 \quad (3)$$

$$\frac{\partial \ln L(x; \theta, \alpha, \beta)}{\partial \alpha} = 0, \quad (4)$$

$$\frac{\partial \ln L(x; \theta, \alpha, \beta)}{\partial \beta} = 0 \quad (5)$$

The solution to equations **Error! Reference source not found.3**, **Error! Reference source not found.4** and **5** are obtained using the *EnvStats* package in R version 4.5.0 (Team, 2025) which provides the MLEs of parameters θ, α , and β .

Return Period and Return Level

Return Level: The annual maximum daily rainfall (AMDR) anticipated on average once every T years is referred to as a T-year return level. The 95% confidence interval for the return level is determined using the Delta method.

Return Period: The average time between the same annual maximum daily rainfall (AMDR) events.

Mean Square Error (MSE)

The average of the squared difference between the observed value and the fitted value.

TABLE 1. Maximum Likelihood Estimates of the Parameters of Fitted Distributions

	Parameters	Scale	Shape	AIC	BIC
	Location				
Gumbel	142.0361	86.5145	---	1228.439	1233.669
	-9.0503	-6.9922			
Weibull	--	218.9618	1.7502	1227.707	1232.938
		-13.171	-0.129		
Generalized EVD	135.5893	80.8326	0.1392	1228.374	1236.219
	-9.8531	-7.8699	-0.1124		

TABLE 2. Kolmogorov-Smirnov Test of Goodness of Fit

Distribution	Test Statistic	P-value
Gumbel	0.099	0.705
Weibull	0.0792	0.9093
Generalized EVD	0.1188	0.47

TABLE 3. Maximum Squared Error

Distribution	Mean Square Error
Gumbel	901.6235
Weibull	925.6437
Generalized EVD	637.9433

FIGURES

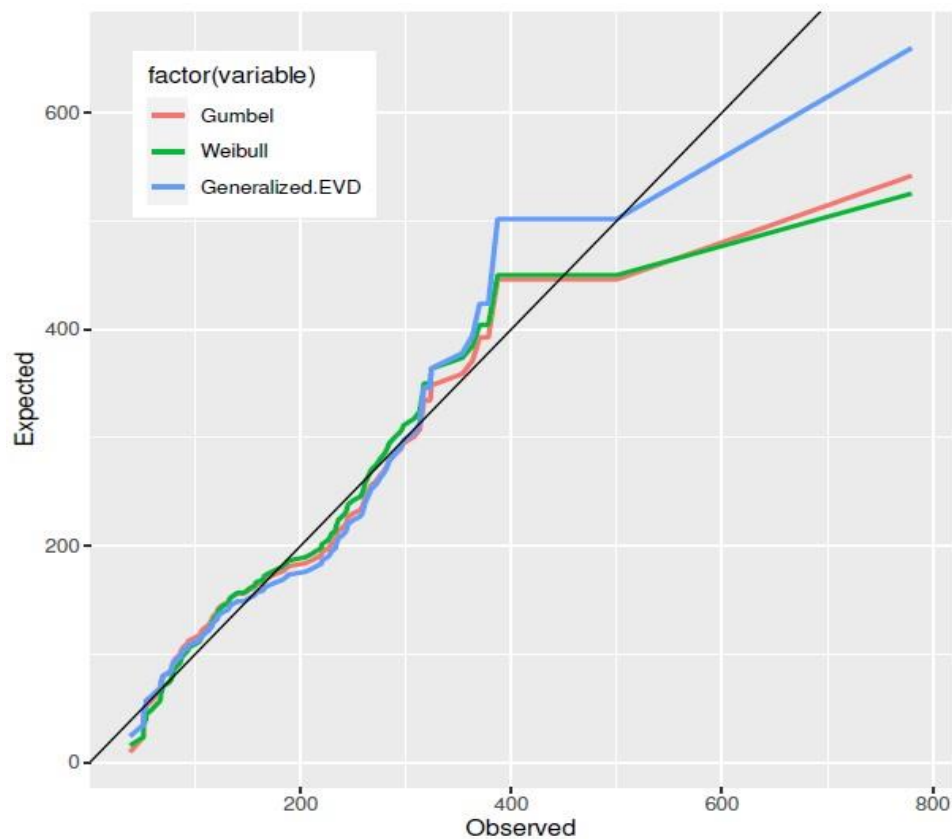


FIGURE 1. Expected AMDR versus Observed AMDR

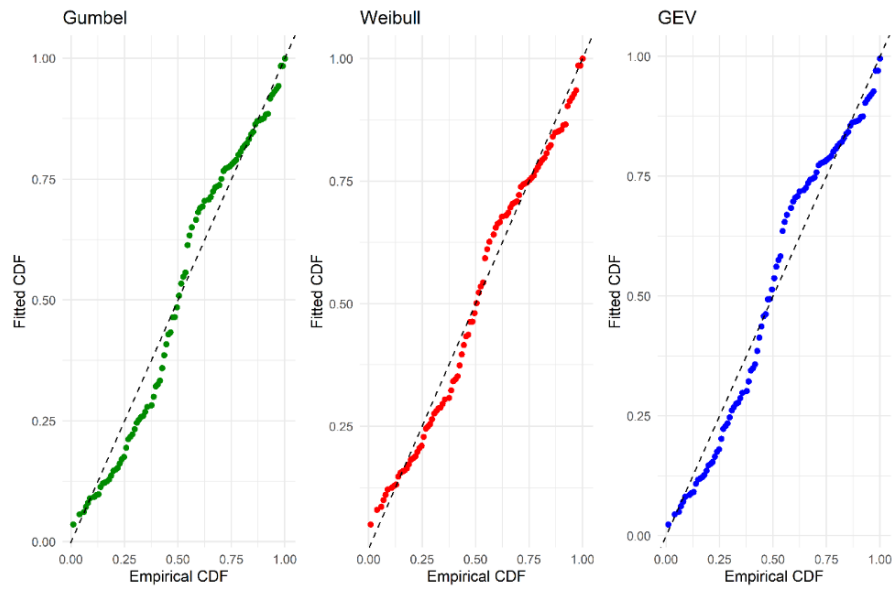


FIGURE 2. Empirical CDF against Fitted CDFs for Gumbel, Weibull, and GEV Distributions

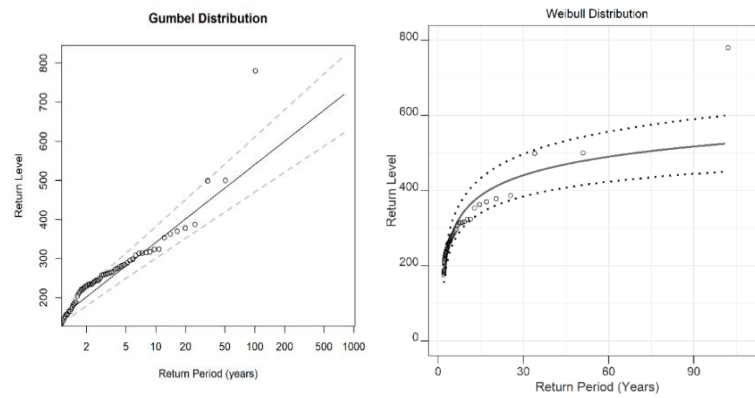


FIGURE 3. (a) Gumbel Distribution ,(b) Weibull Distribution

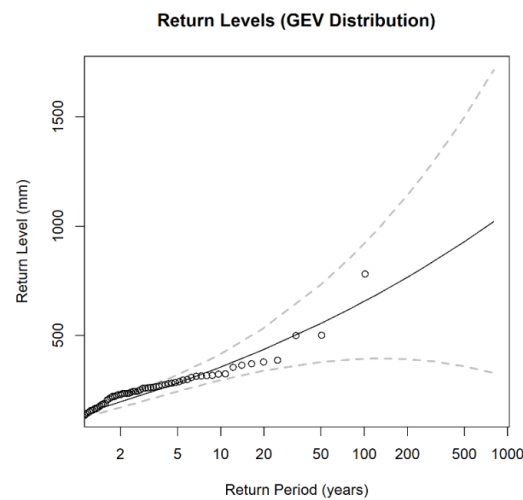


FIGURE 3. (c) Generalized Extreme Value Distribution

Figure 3 Plot of Return Level with 95 % Confidence Interval against Return Period

3. RESULTS AND DISCUSSION

Gumbel, two-parameter Weibull and GEV distributions are fitted to the study variable AMDR. For parameter estimation, the maximum likelihood estimation approach is utilized. In TABLE 1, the estimates of the parameters along with their standard errors (provided in round brackets), are displayed. The values of Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) are also provided in TABLE 1. The Weibull distribution exhibits the minimum value of both AIC and BIC among the three distributions, showing that the Weibull distribution is a good fit. However, the values of AIC and BIC for other distributions, Gumbel and GEV, are close to those of the Weibull distribution, indicating that the fit by the three distributions is not significantly different. Also, the positive value of the shape parameter (0.1392) of the GEV distribution indicates that the distribution has a long right tail, meaning that the extreme events of AMDR are more likely.

TABLE 2 provides the value of the test statistic and p-values of the Kolmogorov–Smirnov (KS) test, since the p-values for each of the fitted distributions are greater than the level of significance 5%, meaning that all the distributions fit well to the data. However, the p-value for the Weibull distribution is highest (0.9093) among all, indicating that it is the best fit among the three, followed by Gumbel (p-value of 0.7050) and GEV (p-value of 0.4750).

TABLE 3 presents the values of the MSE, leading to the conclusion that the minimum MSE value (637.9433) is associated with the GEV distribution, followed by the Gumbel distribution (901.6235), while the Weibull distribution has the highest value (925.6437). This indicates that the GEV is the best-fitting distribution according to this criterion

FIGURE 1 shows the plot of fitted AMDR against observed AMDR. The plot shows that the fitted lines for the GEV distribution are substantially closer to the line through the origin than the fitted lines for the Gumbel and two-parameter Weibull distributions.

FIGURE 2 shows the plot of fitted CDF against an empirical CDF for Gumbel, Weibull and GEV distribution respectively. All the three plots show that there is no significant deviation between the points and line passing through the origin meaning that the Gumbel, Weibull and GEV distributions are well fitted to the observed data. This conclusion is also supported by the KS test presented in TABLE 2.

Figure 3 is a plot of return level with a 95% confidence interval against return period (T). It has three subfigures: Figure 3(a), 3(b), and 3(c) for Gumbel distribution, two-parameter Weibull, and GEV distribution, respectively. The extreme value of 780 mm is included only in the 95% confidence interval of the GEV distribution, however, the CI of Gumbel and two-parameter Weibull distributions do not include this value. The return value for the GEV, Gumbel, and two-parameter Weibull distributions over the return period of 101 years is 659.6849, 541.7374, and 525.2662, respectively. This means that there is a 1/101 chance that there will be 659.6849 mm of precipitation in a year.

4. CONCLUSIONS

This paper presents the study carried out for the modeling of AMDR using Gumbel, two-parameter Weibull, and GEV distributions. The following conclusions were drawn from the study.

- The Gumbel, two-parameter Weibull and GEV distributions provides a good fit at 5% level of significance according to the KS test. However, the highest p-value (0.9093) of Weibull distribution shows that it is best fitted among the three. This claim is also supported by the minimum value of AIC and BIC for two-parameter Weibull distribution.
- Using the mean square error (MSE) criterion, GEV distribution appears to be a good fit for the AMDR data of the Kolhapur district.
- The positive value of the shape parameter of GEV distribution suggests that the distribution is the distribution has a long right tail, meaning that the extreme events of AMDR are more likely.
- The 95% confidence interval of return level for GEV distribution includes the highest observed AMDR value of 780mm, however the 95% CI for Gumbel and two-parameter Weibull distribution do not include this value. Thus, the GEV distribution outperforms others according to this criterion.

The study concludes that the GEV distribution could be used for modeling the AMDR values for the long period data. This could help the administration in management of flood, construction of drainage system, dams etc. across the Kolhapur district.

REFERENCES

- [1]. Bhim Singh, A. V., Deepak Rajpurohit, & Singh, J. (2012). Probability Analysis for Estimation of Annual One Day Maximum Rainfall of Jhalrapatan Area of Rajasthan. *Plant Archives*, 12(2), 1093–1100.
- [2]. Coles, S., Pericchi, L. R., & Sisson, S. (2003). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273(1–4), 35–50. [https://doi.org/10.1016/S0022-1694\(02\)00353-0](https://doi.org/10.1016/S0022-1694(02)00353-0)
- [3]. Ikbal, N. A. M., Halim, S. A., & Ali, N. (2022). Estimating Weibull Parameters Using Maximum Likelihood Estimation and Ordinary Least Squares: Simulation Study and Application on Meteorological Data. *Mathematics and Statistics*, 10(2), 269–292. <https://doi.org/10.13189/ms.2022.100201>
- [4]. IPCC (Ed.). (2023). *Climate Change 2022 - Mitigation of Climate Change: Working Group III Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781009157926>
- [5]. Kumar, R., & Bhardwaj, A. (2015). Probability analysis of return period of daily maximum rainfall in annual data set of Ludhiana, Punjab. *Indian Journal of Agricultural Research*, 49(2), 160. <https://doi.org/10.5958/0976-058x.2015.00023.2>
- [6]. Ndri, K., & Nadarajah, S. (2022). Statistical modeling of monthly maximum rainfall in Senegal. *MAUSAM*, 74(1), 169–176. <https://doi.org/10.54302/mausam.v74i1.5607>
- [7]. Singh, V. (1998). *Entropy-Based Parameter Estimation in Hydrology*. Springer Science.
- [8]. Subudhi, R. (2007). Probability analysis for prediction of annual maximum daily rainfall of Chakapada block of Kandhamal district in Orissa. *Indian Journal of Soil Conservation*, 35(84–85).
- [9]. Team, R. C. (2025). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. <https://www.R-project.org/>
- [10]. Vivekanandan, N. (2012). Intercomparison of Extreme Value Distributions for Estimation of ADMR. *International Journal of Applied Engineering and Technology*, 2(1), 30–37.
- [11]. Vivekanandan, N. (2021). A STUDY ON COMPARISON OF EXTREME RAINFALL ESTIMATES USING L-MOMENTS OF SIX PROBABILITY DISTRIBUTIONS. *Journal of Indian Water Resources Society.*, 41(2), 35–40.