PStatistical Modelling and Best-Fit Distribution Analysis of Annual Rainfall in the Bundelkhand Region, Madhya Pradesh

Abstract

This study evaluates the statistical distribution models most suitable for annual rainfall data across six districts, Chhatarpur, Damoh, Datia, Panna, Sagar, and Tikamgarh, in the Bundelkhand region of Madhya Pradesh. Utilizing a combination of model selection criteria, including Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), along with multiple goodness-of-fit tests (Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling), the research assesses the fit of four probability distributions: Normal, Log-Normal, Gamma, and Weibull. Results indicate that the Log-Normal distribution most accurately models rainfall data in Damoh, Panna, Sagar, and Tikamgarh districts, effectively capturing the characteristic positive skewness of hydrological data. The Gamma distribution is identified as the best fit for the Datia region, while the Weibull distribution provides the best model for Chhatarpur. In all cases, the Normal distribution is consistently outperformed by the other three due to its inability to capture extreme values and data skewness. These insights highlight the importance of selecting suitable statistical models for reliable hydrological analysis, risk assessment, and climate studies in the Bundelkhand region.

Commented [hp1]: 1.Please mention that how many years of rainfall data have been captured?
2.How many number of total observations been considered to apply the distributions?

Keywords: Bundelkhand, Madhya Pradesh, Normal, Log-Normal, Gamma, Weibull.

1. Introduction

Commented [hp2]: 1.Please mention that how many years of rainfall data have been captured?
2.Which season been considered as monsoon season?
3.How many number of observations been considered to apply the distribution?
4. Please include the observations range table for individual region wise rainfall.

Rainfall is one of the most critical meteorological variables influencing the hydrological cycle, agricultural productivity, and water resource management, especially in regions like Bundelkhand, Madhya Pradesh. Accurate analysis of rainfall patterns is essential for effective planning and mitigation of climatic risks, such as droughts and floods, which have significant socioeconomic impacts in semi-arid and monsoon-dependent regions (Sharma, & Singh, 2010; Kumudha, & Kokila, 2023).

The statistical modelling of rainfall data aims to understand both the temporal and spatial variability of precipitation and to characterize its probability distribution for hydrologic design, crop management, and climate studies. Given the nonlinear and positively skewed nature often observed in hydrological datasets, identifying a suitable probability distribution that accurately represents rainfall variability is crucial. Previous studies in India and elsewhere have evaluated various statistical distributions, such as the Normal, Log-Normal, Gamma, and Weibull, for this purpose, finding that no single model universally fits all regions due to local climatic and physiographic conditions (Sharma, & Singh, 2010; Kumudha, & Ramesh, 2023; Nwaigwe *et al*., 2023).

Multiple Indian studies have found the Log-Normal and Gamma distributions are often the most appropriate for modeling annual and monsoon rainfall, although the best fit varies by location and temporal scale. For instance, studies in the North East (Lairenjam, *et al*., 2016), Haryana (Rag, *et al*., 2025), and flood-prone regions of Karnataka (Kumudha, & Ramesh, 2023), have shown Gamma and Log-Normal as leading candidates, with others like Weibull, GEV, and Log Pearson Type III sometimes providing a better fit for particular districts or periods (Kumar, *et al*., 2019; Lairenjam, *et al*., 2016; Kumudha, & Ramesh, 2023; Meena, *et al*., 2019).

The research has highlighted the Log-Normal and Gamma distributions as particularly effective in describing the annual and seasonal rainfall patterns in many parts of India, including flood-prone or semi-arid zones, due to their ability to handle positive skewness and non-negativity. The Weibull distribution, meanwhile, has proven valuable in reliability and life data applications involving skewed hydrological variables. Maximum likelihood estimation and goodness-of-fit tests such as Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling are commonly used to assess how well these models capture the empirical data (Sharma, & Singh, 2010; Kumudha, & Ramesh, 2023).

This study analyzes annual rainfall data from selected districts of the Bundelkhand region by applying these statistical frameworks. By comparing competing distribution models using robust selection criteria (AIC, BIC, and goodness-of-fit tests), we aim to identify the most suitable probability model for each district, thereby enhancing regional hydrological modelling, risk assessment, and long-term climate adaptation strategies.

2. Methodology

Analysis of rainfall data strongly depends on its distribution pattern. It has long been a topic of interest in the fields of meteorology in establishing a probability distribution that provides a good fit to rainfall.

2.1 Normal distribution:

The normal distribution is given by:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

where:

μ: mean

σ: standard deviation

The parameters μ and σ are positive real quantities as is the variable x.

2.2 Gamma distribution (2P):

The gamma distribution is given by:

$$f(x;a,\lambda) = a^{\lambda}x^{\lambda-1}e^{-ax}/\Gamma(\lambda)$$

Where:

a: scale parameter

λ: shape parameter

$\Gamma(\lambda)$: Gamma function

The parameters a and λ are positive real quantities as is the variable x.

2.3 Weibull Distribution (2P):

The Weibull distribution is given by:

$$f(x{:}\,\eta,\sigma) = \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} e^{-\left(\frac{x}{\sigma}\right)^{\eta}},$$

where the variable x and the parameters η and σ are all positive real numbers.

2.4 Log-Normal distribution:

The exponential distribution is given by:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(lnx \quad)^2/2\sigma^2}$$

Where:

μ: Mean of the natural logarithm of the variable

σ: Standard deviation of the natural logarithm of the variable

x: Random variable, must be x>0

Distribution Selection Criteria

Akaike's Information Criterion (AIC)

Akaike first created the AIC in 1973 by adjusting biased empirical information (Akaike, 1973). It is a mathematical technique that is used to assess how well a model matches the data it was created from. AIC is a statistical tool used to analyze various models and identify the best fit for the data (Bevans, 2023). AIC is determined by:

$$\mathrm{AIC} = 2\mathrm{k} - 2\ln(\hat{\mathrm{L}})$$

Bayesian information criterion (BIC)

The BIC is another often-used information criterion. BIC is determined using a Bayesian framework as an estimate of the Bayes factor for two competing models, in contrast to AIC (Schwarz, 1978; Kass and Rafftery, 1995). The BIC is calculated as:

$$\mathrm{BIC} = \mathrm{k} \times \ln \mathrm{N} - 2\ln(\hat{\mathrm{L}})$$

where k = the number of independent variables included in the model's construction, ln = natural log, and $\hat{\mathrm{L}}$ = the maximum likelihood estimates of the model.

2.5 Goodness of Fit

The goodness-of-fit tests, namely, Kolmogorov-Smirnov, Cramér–von Mises and Anderson-Darling, were used at 5% significance level for the selection of the best-fit distribution. The best-fitted distribution is selected based on the minimum error produced, which is evaluated by the following techniques:

2.5.1 Kolmogorov-Smirnov goodness of fit test:

The Kolmogorov-Smirnov (K-S) test is a goodness-of-fit statistic that compares an empirical distribution function (F_x), with a specified distribution function (F_y) (Massey, 1951). Many times test is used as an alternative to the Chi-square goodness-of-fit test. The Kolmogorov-Smirnov statistic (D) can be computed as:

$$D = \max\left|F_x(x) - F_y\right|$$

which measures the distance between the empirical distribution function F_x and the specified distribution function F_y. Obviously, a large difference indicates an inconsistency between the observed data and the statistical models.

2.5.2 Cramér–von Mises goodness of fit test:

The Cramér–von Mises statistic is a widely used goodness-of-fit test that quantifies how well a sample matches a specified cumulative distribution function (CDF). It works by calculating the squared distance between the empirical distribution function (EDF) of the data and the theoretical CDF, thus providing a measure of their overall discrepancy. Unlike some other goodness-of-fit tests, the Cramér–von Mises statistic assigns equal weight to discrepancies across the entire distribution, not just the tails or the center. This makes it a general-purpose

Commented [hp3]: Can this be "centre"?

test ideal for assessing the fit of a model to data without particular sensitivity to any specific region of the distribution (Chiu, & Liu, 2009). The formula is shown below:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(\mathrm{F}(X_i) - \frac{2i-1}{2n} \right)^2$$

Where:

n = sample size

X_i = the i-th order statistic (i.e., sorted sample value)

$\mathrm{F}(x_i)$ = the theoretical cumulative distribution function evaluated at X_i

$\frac{2i-1}{2n}$ = the expected value of the empirical CDF at X_i

2.5.3 Anderson–Darling goodness of fit test:

The Anderson–Darling statistic is a goodness-of-fit test that determines how well a given sample matches a specific cumulative distribution function (CDF). This test is particularly valued for its sensitivity to discrepancies in the tails of the distribution, making it more powerful than some other tests, such as the Kolmogorov–Smirnov test, when tail behavior is of interest. The Anderson–Darling statistic was first introduced by Theodore W. Anderson and Donald A. Darling in (1952). The test calculates a statistic based on the squared difference between the observed and expected CDF values, weighted more heavily in the distribution's tails. This approach allows it to detect deviations that might not be apparent in more central parts of the distribution. The formula is shown below:

$$A^2 = -\mathrm{n} - \frac{1}{n}\sum_{i=1}^{n} [(2i-1)(lnF(X_i) + \ln\,(1 - F(X_{(n+1-i)}))]$$

Where:

n= Number of observations

X_i= The i-th order statistic (sorted sample)

$F(X_i)$= The theoretical CDF of the distribution you're testing against (e.g. normal, exponential)

$F(X_{(n+1-i)})$= The CDF evaluated at the complementary ordered value

3. Results and discussion

The Bundelkhand region is a historically and geographically important area in central India, spanning primarily across northern Madhya Pradesh and lower Uttar Pradesh. The larger part of Bundelkhand lies within Madhya Pradesh, specifically covering key districts such as Chhatarpur, Tikamgarh, Damoh, Sagar, Datia, and Panna (Jain, 2002). This region forms a distinct plateau, recognized for its semi-arid climate, characterized by low to moderate rainfall, hot summers, and dry conditions outside the monsoon period (Scott and Divakar, 2024).

Geographically, Bundelkhand sits between the fertile Gangetic plains to the north and the rugged Vindhyan highlands to the south, making it a transitional zone between two significant landscapes. It is roughly positioned between 23°20′N to 26°20′N latitude and 78°20′E to 81°40′E longitude, reinforcing its identity as a bridge between plains and uplands.

Agriculture dominates the regional economy, but farming is mostly rain-fed, making communities highly vulnerable to erratic monsoon rains and recurring droughts. As a result, crop failures are a persistent challenge affecting rural livelihoods.

Despite climatic and economic challenges, Bundelkhand boasts several major rivers: Sindh, Betwa, Ken, Bagahin, Tons, Pahuj, Dhasan, and Chambal. These rivers are crucial lifelines, supporting drinking water supplies, irrigation, and domestic needs not only for Bundelkhand but also for neighboring areas in Uttar Pradesh.

Overall, Bundelkhand's landscape, water resources, and economic dependence on rain-fed agriculture play a pivotal role in shaping the lives and environment of its population. The base map of the study area is shown in Figure 1.

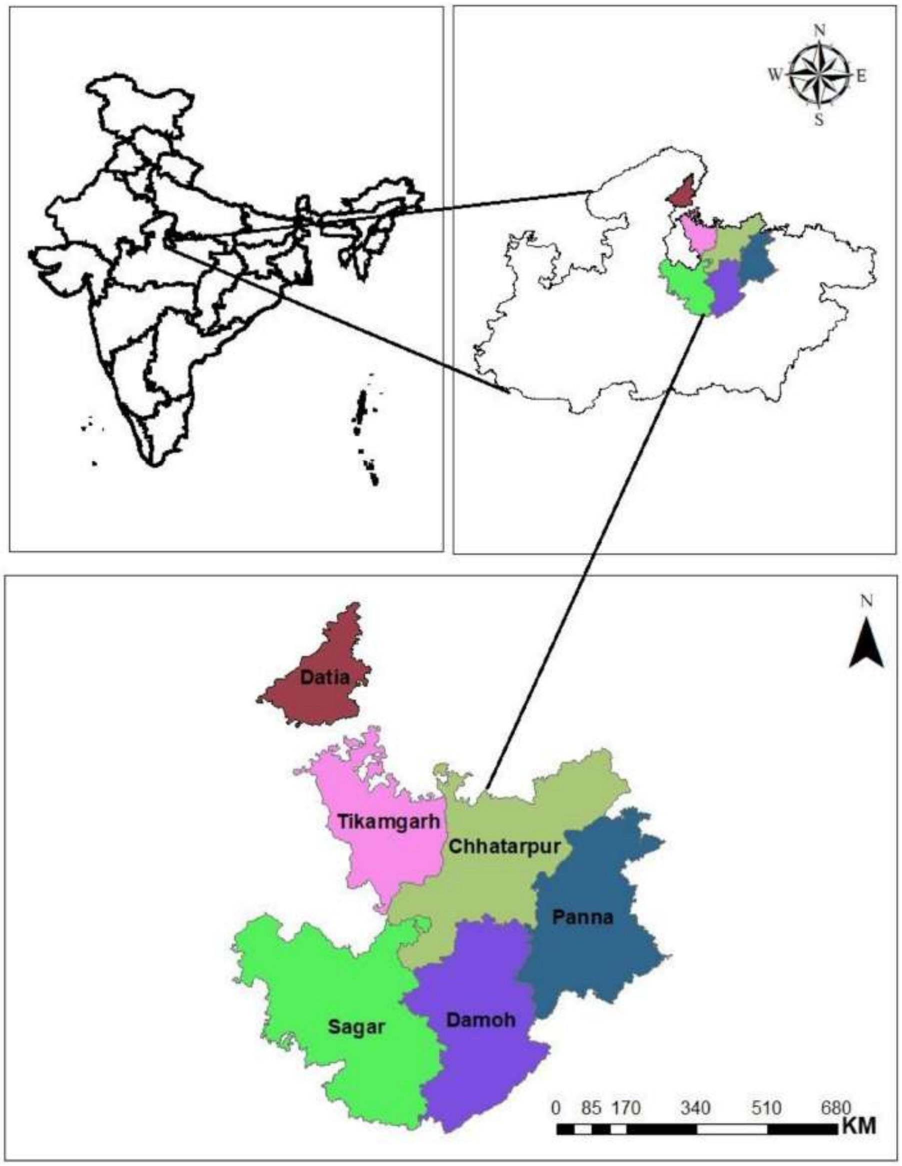


Figure 1. Map of the Study Area

Table 1: The fit distribution resulting from the Chhatarpur region's annual rainfall forms the basis for the AIC and BIC selection criteria.

Best-Fit Distribution	AIC	BIC
Gamma	202.0834	203.4995
Weibull	**201.8482**	**203.2643**
Log-normal	202.4348	203.8509
Normal	201.8569	203.2830

Above Table 1, shows that, based on model selection criteria, including AIC and BIC, the Weibull distribution provides the best fit for the Chhatarpur region's annual rainfall data, as it has the lowest AIC (201.85) and BIC (203.26), closely followed by the Normal distribution. This suggests that the rainfall pattern follows a moderately skewed distribution, with the Weibull distribution being most appropriate for modeling rainfall behavior in this case.

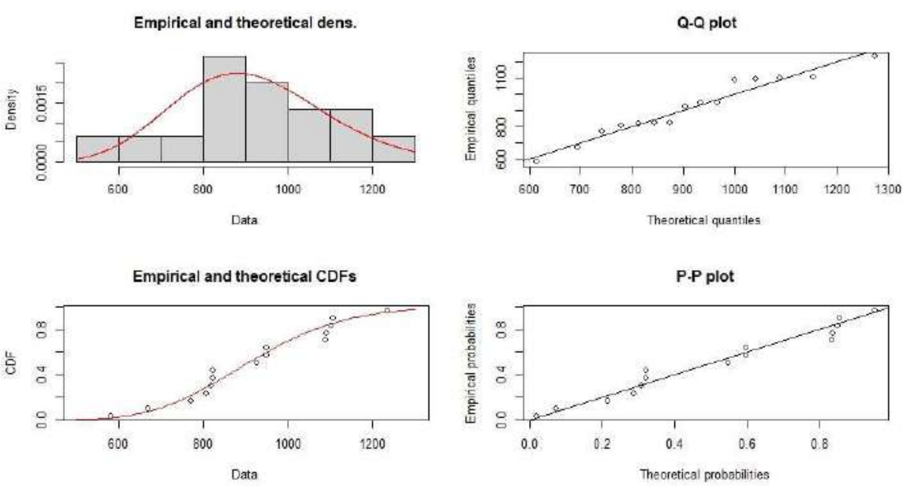


Figure 2: Chhatarpur region's Gamma distribution density, Q-Q, CDF, P-P, and plots.

Commented [hp4]: Should this ", and" be replaced before "P-P"?
For example: "**Q-Q, CDF, and P-P plots**".

Figure 2, The Gamma distribution provides a reasonably good fit for the Chhatarpur region's annual rainfall data, capturing the positive skewness more effectively than the normal distribution, as reflected by the alignment of the empirical data with the theoretical model in the density, Q-Q, CDF, and P-P plots.

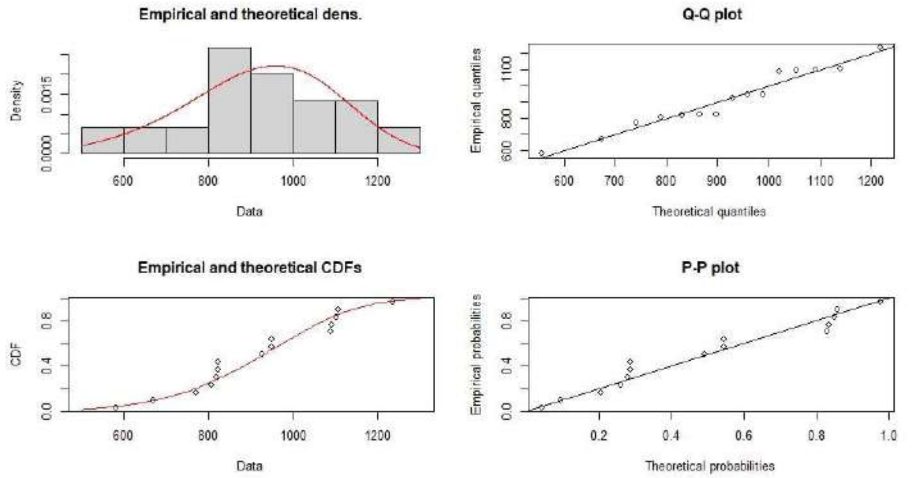
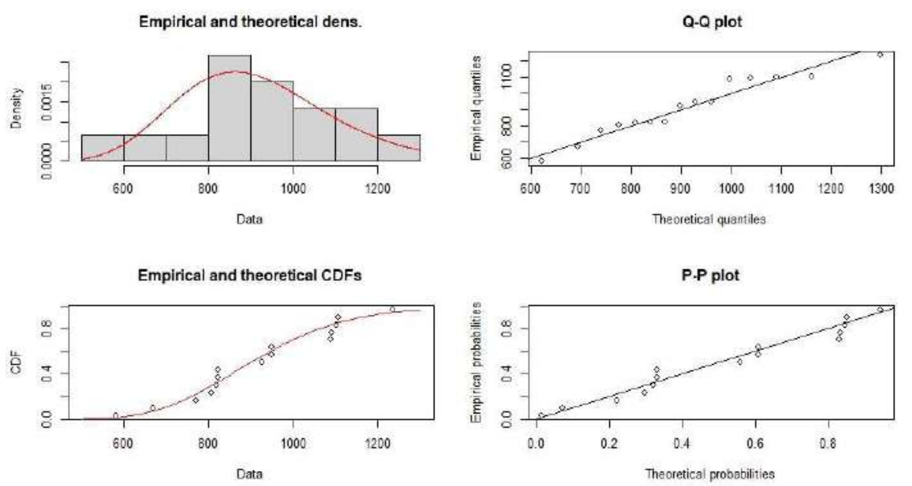


Figure 3: Chhatarpur region's Weibull distribution density, Q-Q, CDF, P-P, and plots.

Figure 3, The Weibull distribution provides an acceptable model for the Chhatarpur region's annual rainfall data, especially useful for reliability and life data where skewness exists. However, compared to distributions like Gamma, Log-Normal and Normal (based on your earlier images), the Weibull shows slightly more deviation at the tails, making it a slightly less optimal fit in those cases.



Commented [hp5]: 1.The x-axis for Q-Q plot and CDF plot should include a range start from 500, instead of 600, as there is one value which is less than 600 in the graph.
2.The x-axis of P-P plot should start the range from 0.0 as similar to Figure 2.
3. Please clarify the term "Data" in x-axis label.

Commented [hp6]: Should this ", and" be replaced before "P-P"?
For example: "**Q-Q, CDF, and P-P plots**".

Figure 4: Chhatarpur region's Log-normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 4, The Log Normal distribution is a statistically appropriate and well-fitting model for the Chhatarpur region's annual rainfall data, effectively representing its positively skewed nature with strong alignment across density, Q–Q, CDF, and P–P plots.

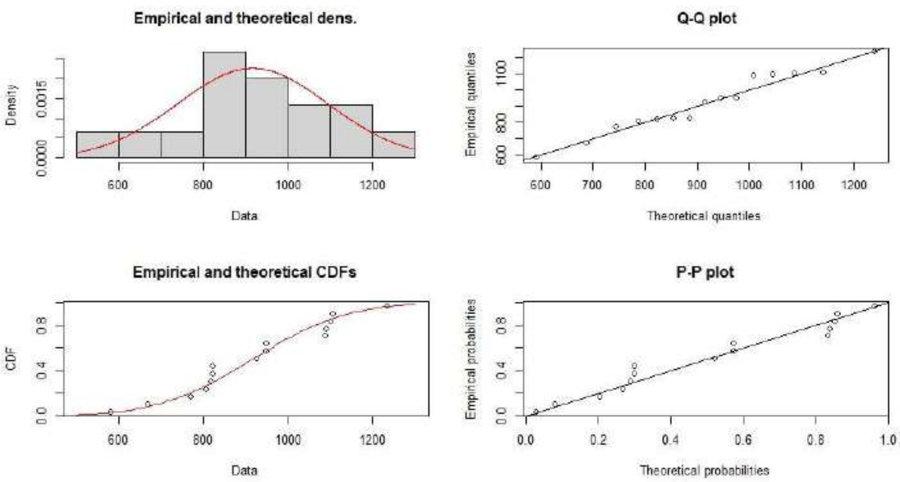


Figure 5: Chhatarpur region's Normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 5, The Normal distribution is a reasonable approximation for the Chhatarpur region's annual rainfall data, but it may not fully capture extreme low or high rainfall values. This indicates light deviation from normality, especially in the tails. Consider testing Gamma, Weibull, or Log-Normal distributions for a potentially better fit.

Goodness-of-fit statistics

Table 2: Comparison of Goodness-of-Fit Tests for Normal, Log-Normal, Gamma, and Weibull Distributions of annual rainfall data in the Chhatarpur region.

Statistic	Normal	Log-Normal	Gamma	Weibull
Kolmogorov–Smirnov	0.1698	0.1649	0.1672	**0.1590**
Cramér–von Mises	0.0612	0.0567	0.0566	**0.0534**

Anderson–Darling	0.3625	0.3684	0.3556	**0.3228**

Above Table 2), shows that to analyse the annual rainfall data, several probability distributions, Normal, Log-Normal, Gamma, and Weibull, were fitted to the annual rainfall data, and their fit was evaluated using Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling statistics. These tests measure the agreement between the observed data and each theoretical distribution, with smaller values indicating better fit. In this analysis, the Weibull distribution shows the lowest values for all three tests, suggesting it provides the best fit to the annual rainfall data compared to the other distributions.

Table 3: The fit distribution resulting from the Damoh region's annual rainfall forms the basis for the AIC and BIC selection criteria.

Best-Fit Distribution	AIC	BIC
Gamma	212.5108	213.9269
Weibull	214.4215	215.8376
Log-normal	**212.2702**	**213.6863**
Normal	213.5341	214.9502

Above Table 3, showed that, the based on model selection criteria, including AIC and BIC the lowest AIC (212.27) and BIC (213.69) values, the log-normal distribution provides the best fit for the Damoh region's annual rainfall data among the tested distributions (Gamma, Weibull, Normal), indicating a positively skewed rainfall pattern typical of hydrological data, with Gamma as a close alternative.

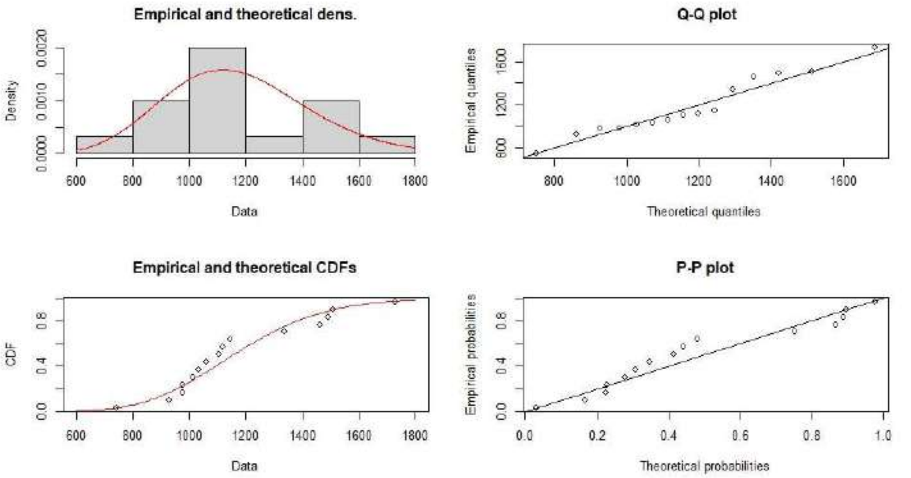


Figure 6: Damoh region's Gamma distribution density, Q-Q, CDF, P-P, and plots.

Figure 6, the Gamma distribution is a statistically appropriate and well-fitting model for the annual rainfall data in the Damoh region, effectively representing its skewed nature with strong alignment across density, Q–Q, CDF, and P–P plots.

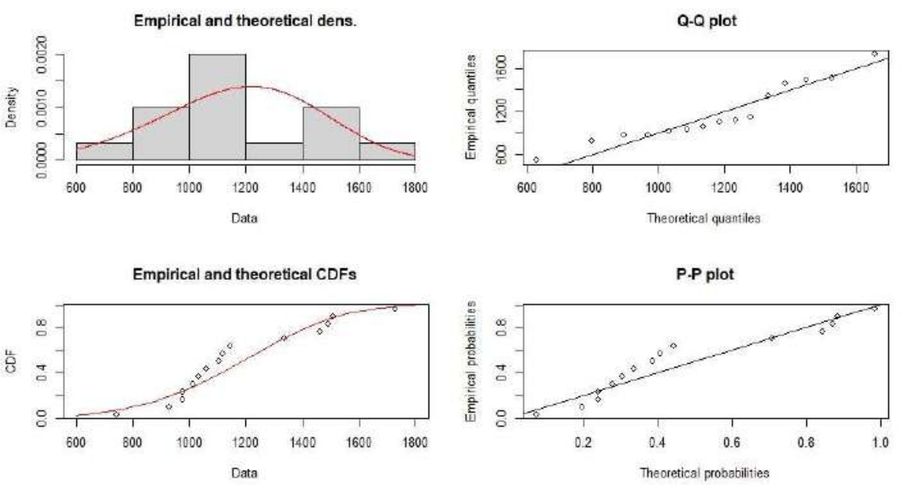


Figure 7: Damoh region's Weibull distribution density, Q-Q, CDF, P-P, and plots.

Figure 7, the Weibull distribution provides a reasonably good fit for the Damoh region's annual rainfall data, effectively capturing the central trend with moderate deviations at the tails, as supported by the density, Q–Q, CDF, and P–P plots.

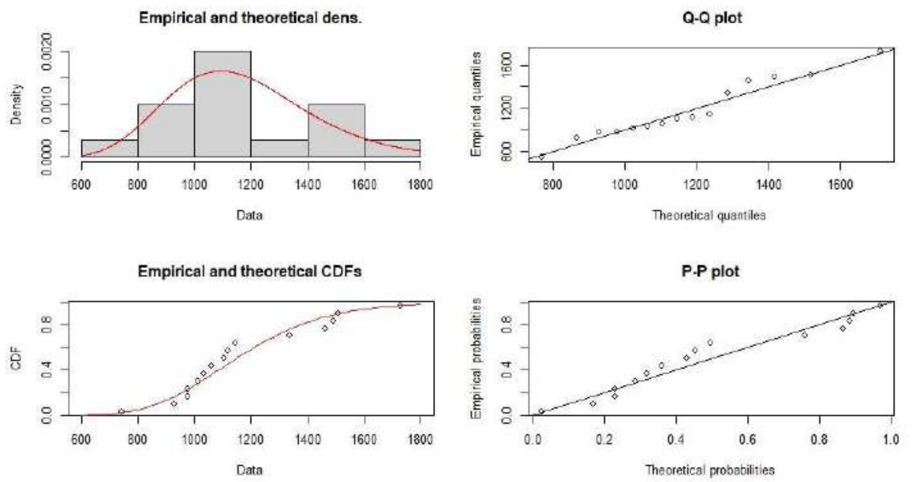


Figure 8: Damoh region's Log-normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 8, The log-normal distribution provides the best fit for the Damoh region's annual rainfall data, accurately capturing its skewed nature and showing the closest alignment across all diagnostic plots compared to the gamma, Weibull, and normal distributions.

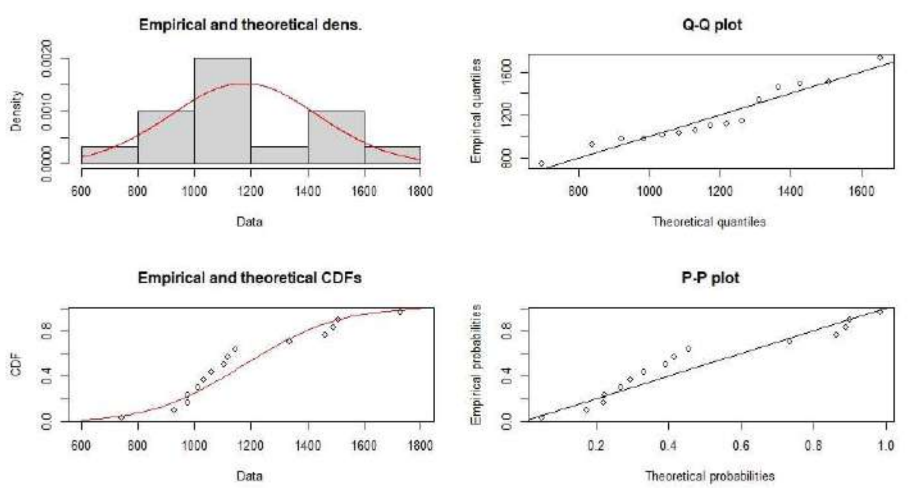


Figure 9: Damoh region's Normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 9, The normal distribution does not provide a good fit for the annual rainfall data of the Damoh region, as the data exhibit skewness that the symmetric normal distribution fails to capture, as indicated by deviations in the Q-Q and P-P plots and a less accurate match in the density and CDF curves.

Goodness-of-fit statistics

Table 4: Comparison of Goodness-of-Fit Tests for Normal, Log-Normal, Gamma, and Weibull Distributions of annual rainfall data in the Damoh region.

Statistic	Normal	Log-Normal	Gamma	Weibull
Kolmogorov–Smirnov	0.2130	**0.1706**	0.1853	0.2235
Cramér–von Mises	0.1124	**0.0751**	0.0867	0.1178
Anderson–Darling	0.5846	**0.4183**	0.4645	0.6169

Above Table 4, to analyze the annual rainfall data, several probability distributions—Normal, Log-Normal, Gamma, and Weibull were fitted to the observed values. Goodness-of-fit tests including Kolmogorov Smirnov, Cramér von Mises, and Anderson Darling were applied to evaluate how well each distribution models the rainfall patterns. These tests measure the differences between the empirical data and the theoretical distributions, with lower values indicating a better fit. Based on the results, the **Log-Normal** distribution consistently shows the lowest values across all tests, suggesting it best captures the characteristics of the annual rainfall data.

Table 5: The fit distribution resulting from the Datia region's annual rainfall forms the basis for the AIC and BIC selection criteria.

Best-Fit Distribution	AIC	BIC
Gamma	**198.2567**	**199.1728**
Weibull	198.4489	199.865
Log-normal	199.3122	200.7283
Normal	198.4408	199.8569

Above Table 5, it is shown that the Gamma distribution provides the best fit for the Datia region's annual rainfall data, both visually and statistically. It effectively captures the positively skewed nature of rainfall, and aligns well across all diagnostic plots (density, Q–Q, P–P, and CDF).

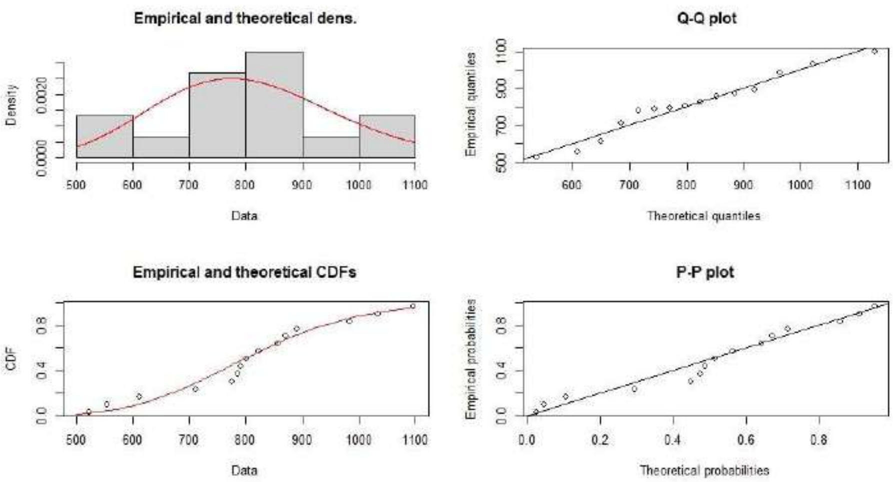


Figure 10: Datia region's Gamma distribution density, Q-Q, CDF, P-P, and plots.

Figure 10, the Gamma distribution provides a superior fit to the Datia region's annual rainfall data compared to the Normal distribution. It better captures the right-skewed nature and non-negative values typical of rainfall. Diagnostic plots (Q-Q, P-P, CDF, and density) confirm that the Gamma distribution models both the central tendency and tail behavior more accurately.

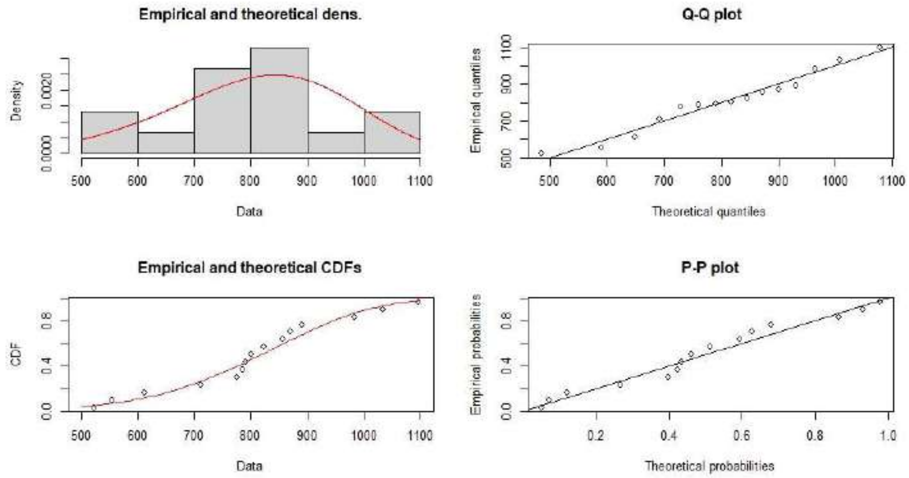


Figure 11: Datia region's Weibull distribution density, Q-Q, CDF, P-P, and plots.

Figure 11, the Weibull distribution provides a good fit to the Datia region's annual rainfall data. It models right-skewed and non-negative data effectively, which is appropriate for rainfall analysis.

All diagnostic plots (density, Q-Q, CDF, and P-P) show good agreement between empirical data and the Weibull model, with only minor deviations in the tails.

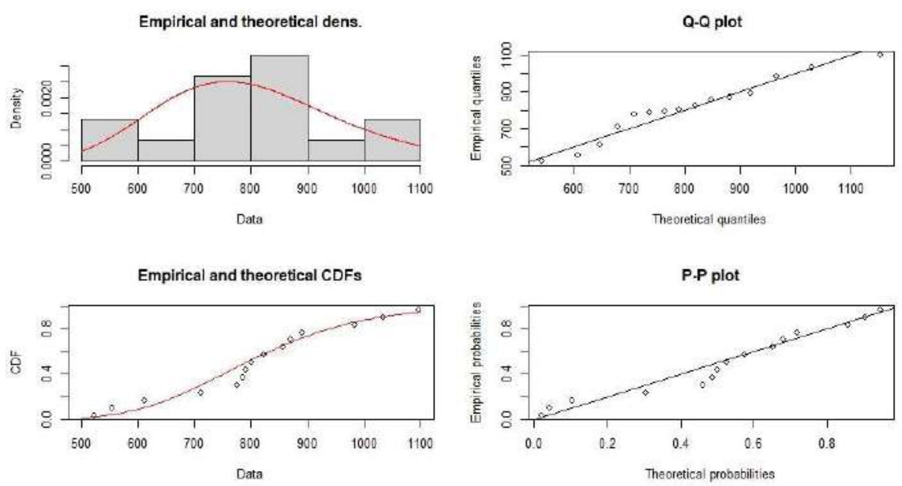


Figure 12: Datia region's Log-normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 12, the Log-Normal distribution fits the Datia region's annual rainfall data very well, especially capturing the positive skewness and non-negative nature of the data. All four plots density, Q-Q, CDF, and P-P, confirm that the log-normal distribution accurately represents both the central tendency and tail behavior of the data.

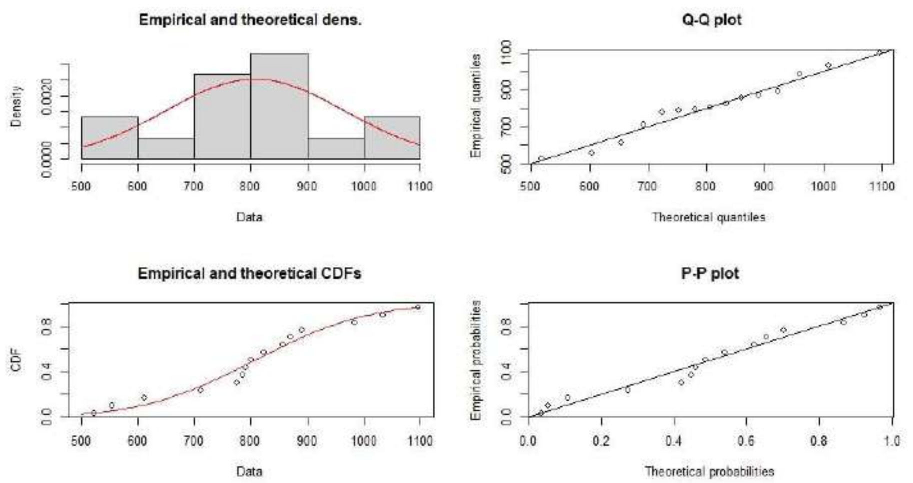


Figure 13: Datia region's Normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 4.3(l), the Normal distribution provides a moderate fit to the Datia region's annual rainfall data. It captures the central peak well, but fails to capture the skewness and tail behavior, which are typical characteristics of rainfall data.

Goodness-of-fit statistics

Table 6: Comparison of Goodness-of-Fit Tests for Normal, Log-Normal, Gamma, and Weibull Distributions of annual rainfall data in the Datia region.

Statistic	Normal	Log-Normal	Gamma	Weibull
Kolmogorov–Smirnov	0.1533	0.1947	**0.1209**	0.1316
Cramér–von Mises	0.0430	0.0677	**0.0268**	0.0429
Anderson–Darling	0.2692	0.4013	**0.2422**	0.2637

Above Table 6, shows that the annual rainfall data was fitted to four distributions, Normal, Log-Normal, Gamma, and Weibull, and evaluated using Kolmogorov–Smirnov, Cramér–von Mises, and Anderson-Darling tests to assess the quality of fit. Lower values of these statistics indicate a better match between the distribution and the observed data. The Gamma distribution demonstrates the smallest values across all three tests, indicating it provides the best fit among the considered models for the rainfall data.

Panna region

Table 7: The fit distribution resulting from the Panna region's annual rainfall forms the basis for the AIC and BIC selection criteria.

Best-Fit Distribution	AIC	BIC
Gamma	206.1525	207.5686
Weibull	209.2943	210.7104
Log-normal	**205.6083**	**207.0244**
Normal	207.6667	209.0828

Above Table 7 showed that, based on the lowest AIC (205.61) and BIC (207.02) values, the log-normal distribution is identified as the best-fit model for the Panna region's annual rainfall data. This suggests that the data is positively skewed, which aligns with common characteristics of rainfall patterns. The Gamma and Normal distributions are close contenders, but their higher AIC and BIC values indicate a slightly poorer fit compared to the log-normal model.

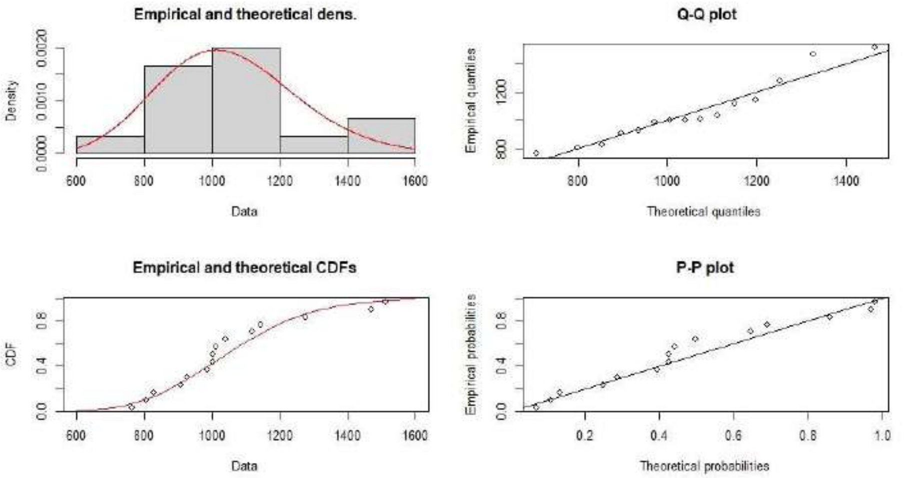


Figure 14: Panna region's Gamma distribution density, Q-Q, CDF, P-P, and plots.

Figure 14, the diagnostic plots indicate that the Gamma distribution is a good fit for modelling the Panna region's annual rainfall data. It effectively captures the positive skewness, and the agreement between empirical and theoretical distributions is visually evident across all plots.

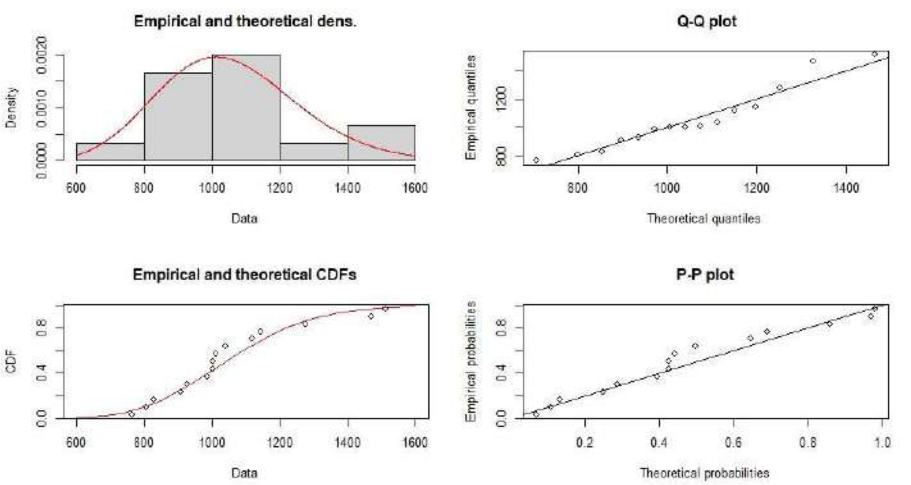


Figure 15: Panna region's Weibull distribution density, Q-Q, CDF, P-P, and plots.

Figure 15, the Weibull distribution provides a fair fit to the Panna region's annual rainfall data. It handles the skewness and shape moderately well and matches the cumulative and probabilistic structure reasonably, though not perfectly. Minor deviations in Q-Q and P-P plots suggest that it may not capture extreme or central tendencies as well as the Gamma distribution, but it is still a valid candidate model.

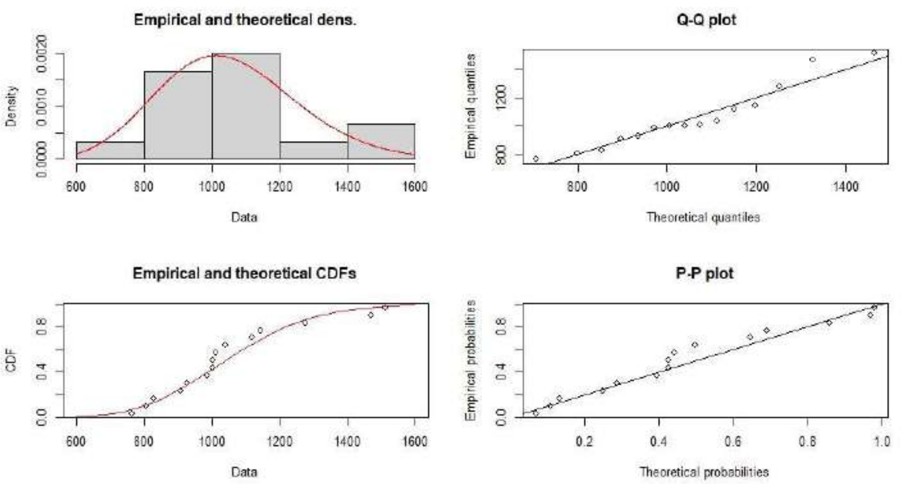


Figure 16: Panna region's Log-normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 16, The log-normal distribution provides a very good fit to the Panna region's annual rainfall data. It successfully captures the right skewness, fits the density, and aligns well with the quantiles and cumulative probabilities. While there may be minor discrepancies at the extremes, the model effectively represents the central and majority behavior of the dataset.

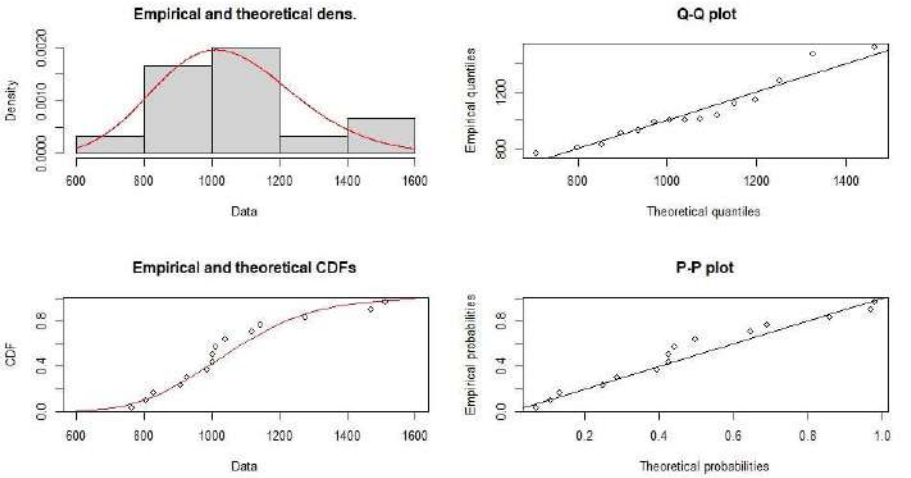


Figure 17: Panna region's normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 17, The normal distribution provides a reasonable fit to the central part of the Panna region's annual rainfall data, but it underestimates tail behavior, particularly the right-skewed extreme values. It is not ideal for modeling rainfall or other naturally skewed data where the gamma or log-normal might perform better.

Goodness-of-fit statistics

Table 8: Comparison of Goodness-of-Fit Tests for Normal, Log-Normal, Gamma, and Weibull Distributions of annual rainfall data in the Panna region.

Statistic	Normal	Log-Normal	Gamma	Weibull
Kolmogorov–Smirnov	0.1931	**0.1547**	0.1684	0.2029
Cramér–von Mises	0.0901	**0.0523**	0.0632	0.1153
Anderson–Darling	0.5707	**0.3449**	0.4097	0.6986

Above Table 8, shows that the goodness-of-fit tests for the annual rainfall data were performed on four distributions: Normal, Log-Normal, Gamma, and Weibull. The Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling statistics were used to evaluate how well each distribution fits the data, with lower values indicating a better fit. Among the tested models, the

Log-Normal distribution consistently shows the lowest values across all three tests, suggesting it provides the best representation of the annual rainfall data compared to the Normal, Gamma, and Weibull distributions.

Table 9: The fit distribution resulting from the Sagar region's annual rainfall forms the basis for the AIC and BIC selection criteria.

Best-Fit Distribution	AIC	BIC
Gamma	214.6716	216.0877
Weibull	216.4558	217.8719
Log-normal	**214.2583**	**215.6744**
Normal	215.9553	217.3714

Above Table 9, shows that the log-normal distribution provides the best fit for the Sagar region's annual rainfall data, with the lowest AIC (214.26) and BIC (215.67) values among all tested distributions. This indicates that the rainfall data is positively skewed and better modelled by a log-normal distribution, which is commonly suited for hydrological data. The Gamma distribution is a close second, while Weibull and Normal show comparatively poorer fits.

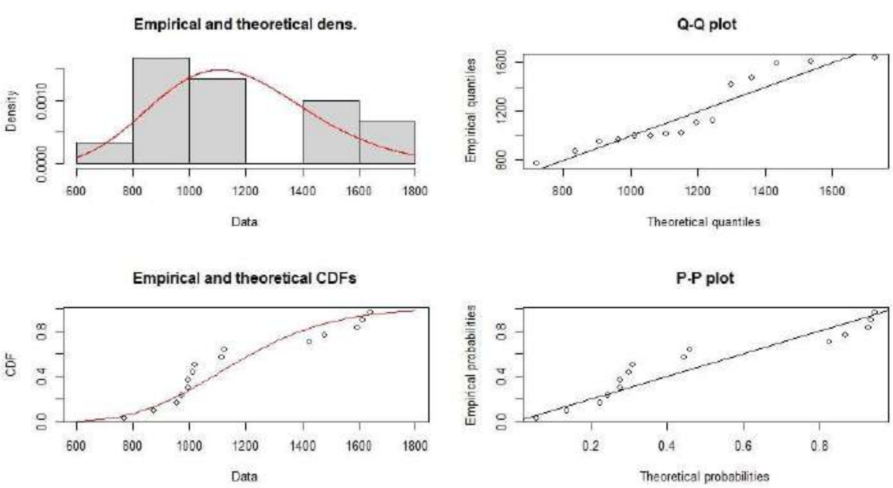


Figure 18: Sagar region's Gamma distribution density, Q-Q, CDF, P-P, and plots.

Figure 18 The Gamma distribution provides a reasonably good fit for the Sagar region's annual rainfall data, capturing the positive skewness more effectively than the normal distribution, as reflected by the alignment of the empirical data with the theoretical model in the density, Q-Q, CDF, and P-P plots.

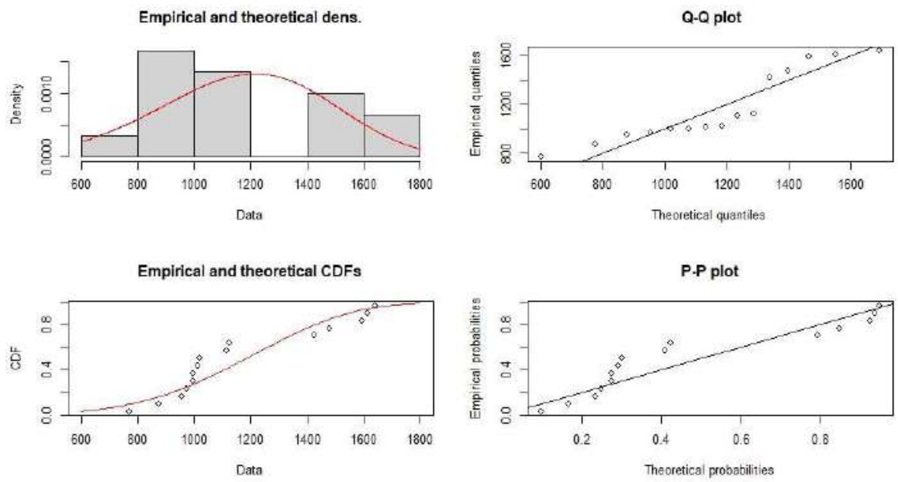


Figure 19: Sagar region's Weibull distribution density, Q-Q, CDF, P-P, and plots.

Figure 19, the Weibull distribution provides a good fit to the Sagar region's annual rainfall data. It models right-skewed and non-negative data effectively, which is appropriate for rainfall analysis.

All diagnostic plots (density, Q-Q, CDF, and P-P) show good agreement between empirical data and the Weibull model, with only minor deviations in the tails.

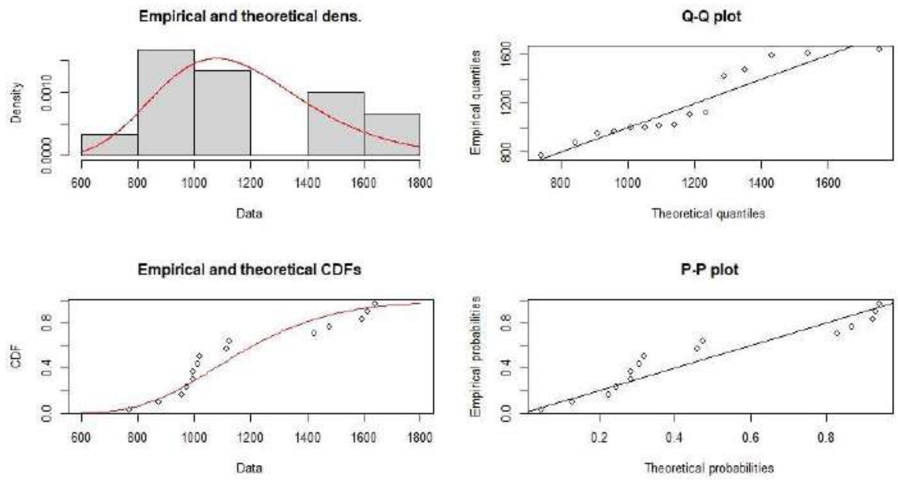


Figure 20: Sagar region's Log-normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 20, the Log-Normal distribution fits the Sagar region's annual rainfall data very well, especially capturing the positive skewness and non-negative nature of the data. All four plots density, Q-Q, CDF, and P-P confirm that the log-normal distribution accurately represents both the central tendency and tail behavior of the data.

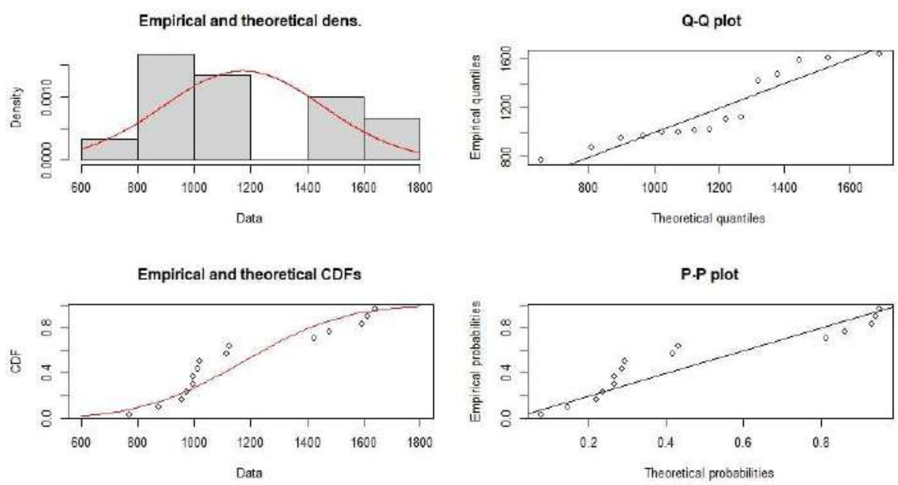


Figure 21: Sagar region's normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 21, the normal distribution provides a reasonable fit to the central part of the data, but it underestimates tail behavior, particularly the right-skewed extreme values. It is not ideal for modeling rainfall or other naturally skewed data where the gamma or log-normal might perform better.

Goodness-of-fit statistics

Table 10: Comparison of Goodness-of-Fit Tests for Normal, Log-Normal, Gamma, and Weibull Distributions of annual rainfall data in the Sagar region.

Statistic	Normal	Log-Normal	Gamma	Weibull
Kolmogorov-Smirnov	0.2380	**0.2153**	0.2238	0.2445
Cramer-von Mises	0.1835	**0.1392**	0.1537	0.1820
Anderson-Darling	0.9958	**0.7729**	0.8428	0.9778

Table 10, shows the fit of four distributions, Normal, Log-Normal, Gamma, and Weibull to the annual rainfall data was evaluated using Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling statistics. Lower values indicate a better fit to the observed data. Among the tested distributions, the Log-Normal distribution consistently shows the lowest goodness-of-fit statistics, suggesting it is the most appropriate model for describing the rainfall data compared to the Normal, Gamma, and Weibull distributions.

Table 11: The fit distribution resulting from the Tikamgarh region's annual rainfall forms the basis for the AIC and BIC selection criteria.

Best-Fit Distribution	AIC	BIC
Gamma	212.4874	213.9035
Weibull	214.171	215.5871
Log-normal	**212.2368**	**213.6529**
Normal	213.636	215.0521

Table 11, shows that of the four distributions tested, the log-normal distribution shows the lowest AIC (212.24) and BIC (213.65), indicating that it provides the best statistical fit to the

Tikamgarh region's annual rainfall data. This implies that the rainfall values are positively skewed, and the log-normal distribution effectively captures this characteristic. The Gamma distribution is a close second, while Weibull and Normal are comparatively less suitable for modeling this dataset.

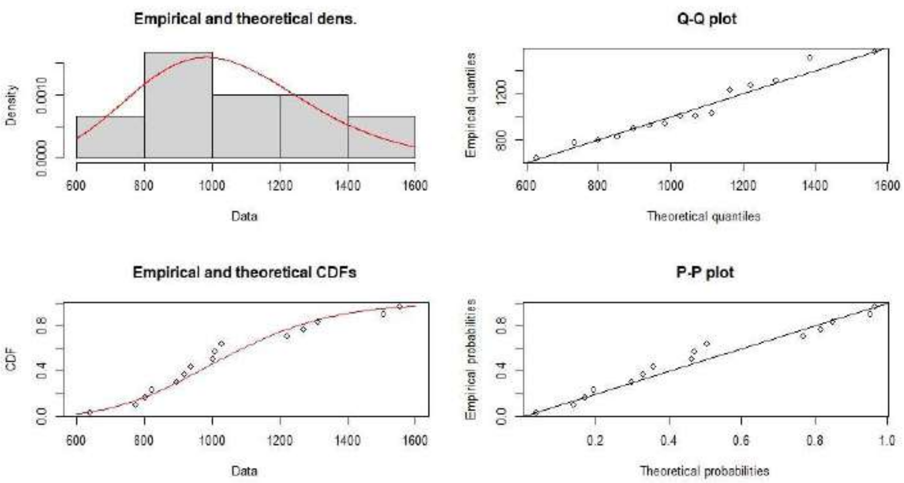


Figure 22: Tikamgarh region's Gamma distribution density, Q-Q, CDF, P-P, and plots.

Figure 22, the Gamma distribution provides a reasonably good fit for the Tikamgarh region's annual rainfall data, capturing the positive skewness more effectively than the normal distribution, as reflected by the alignment of the empirical data with the theoretical model in the density, Q-Q, CDF, and P-P plots.

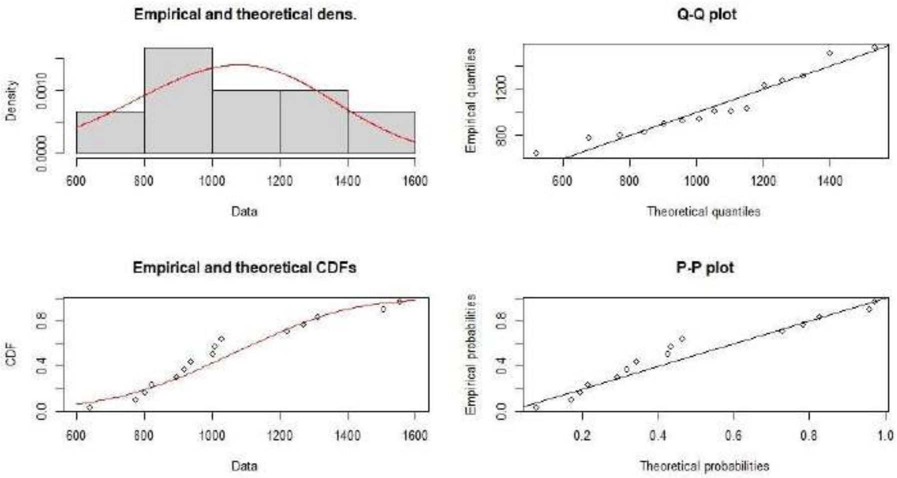


Figure 23: Tikamgarh region's Weibull distribution density, Q-Q, CDF, P-P, and plots.

Figure 23, the Tikamgarh region's annual rainfall data fits well with the Weibull distribution. It accurately models non-negative and right-skewed data, which is suitable for rainfall analysis. With just slight variations in the tails, all diagnostic plots (density, Q-Q, CDF, and P-P) demonstrate strong agreement between empirical data and the Weibull model.

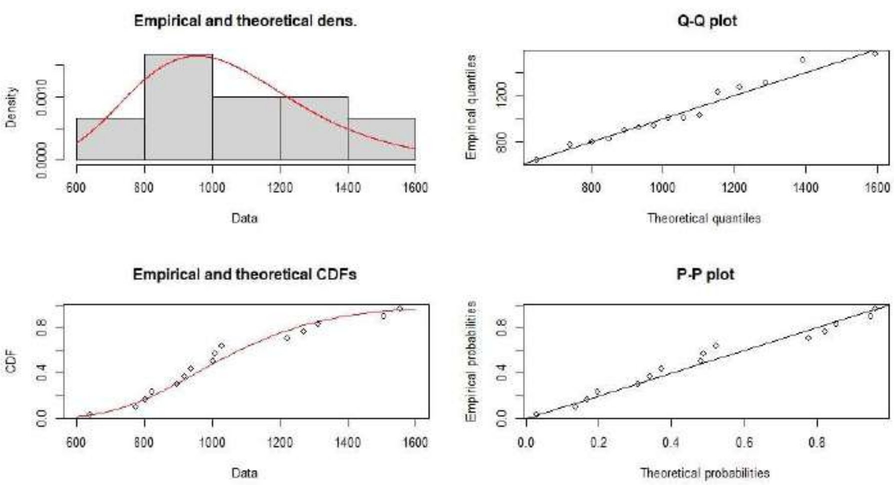


Figure 24: Tikamgarh region's Log-normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 24, the Tikamgarh region's annual rainfall data is well-fitted by the Log-Normal distribution, which particularly captures the data's positive skewness and non-negative character. The log-normal distribution correctly depicts the data's central tendency and tail behavior, according to the four plots of density, Q-Q, CDF, and P-P.

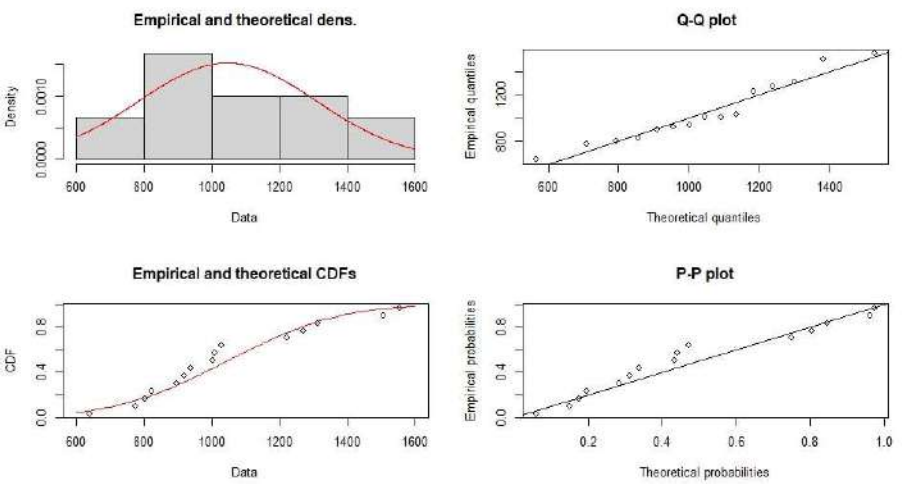


Figure 25: Tikamgarh region's Normal distribution density, Q-Q, CDF, P-P, and plots.

Figure 25, the normal distribution fits the data reasonably well in the middle, it underrepresents the tail behaviour, especially the extreme values that are tilted to the right. The gamma or log-normal may work better when modelling rainfall or other naturally skewed data; thus, it is not the best option.

Goodness-of-fit statistics

Table 12: Comparison of Goodness-of-Fit Tests for Normal, Log-Normal, Gamma, and Weibull Distributions of annual rainfall data in the Tikamgarh region.

Statistic	**Normal**	**Log-Normal**	**Gamma**	**Weibull**
Kolmogorov–Smirnov	0.1943	**0.1456**	0.1623	0.2035
Cramér–von Mises	0.0768	**0.0429**	0.0524	0.0806
Anderson–Darling	0.4361	**0.2590**	0.3041	0.4614

Table 12 shows that the annual rainfall data was fitted to four distributions, Normal, Log-Normal, Gamma, and Weibull and their fit was evaluated using Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling statistics. Lower values indicate a better fit between the model and the observed data. Among these, the Log-Normal distribution consistently has the lowest statistics across all tests, suggesting it provides the best fit for modeling the annual rainfall compared to the Normal, Gamma, and Weibull distributions.

4. Calculation

The analysis involved collecting annual rainfall data for each district and fitting four candidate probability distributions, Normal, Log-Normal, Gamma, and Weibull, using maximum likelihood estimation to determine parameter values. Model selection criteria, including Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), were computed for each fitted distribution, with the lowest values indicating the best balance between goodness of fit and model complexity. Additionally, goodness-of-fit tests, Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling, were applied to assess the agreement between empirical data and theoretical distributions, where smaller statistics signify better fit. Visual diagnostic plots (density, Q-Q, CDF, and P-P) further confirmed the fitting adequacy. The combined evaluation of these numeric criteria and graphical analysis enabled the identification of the most appropriate probability distribution for modeling annual rainfall in each district of Bundelkhand.

COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

References

Commented [hp7]: Please add 4-5 more recent references from the similar area of interest.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255-265.

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 193-212.

Bevan, P., Ferreira, G. B., Ingram, D. J., Rowcliffe, M., Young, L., Freeman, R., & Jones, K. E. (2023). Biodiversity shows unique responses to land-use change across regional biomes. *bioRxiv*, 2023-03.

Chiu, S. N., & Liu, K. I. (2009). Generalized Cramér–von Mises goodness-of-fit tests for multivariate distributions. *Computational Statistics & Data Analysis*, *53*(11), 3817-3834.

Jain, R. K. (2002). *Between history and legend: Status and power in Bundelkhand*. Orient Blackswan.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*(430), 773-795.

Kumar, K. P., Swain, D. K., & Vinay, T. V. (2019). Characterization of Rainfall through Probability Distributions for Yadgir District in Karnataka, India. *Int. J. Curr. Microbiol. App. Sci*, *8*(4), 1099-1113.

Kumudha, H. R., & Kokila R., (2023). Statistical Analysis and Distribution Modeling of Rainfall Data in Flood-Prone Regions of Karnataka State, *Tuijin Jishu/Journal of Propulsion Technology*, *44*(4), 6848-6862.

Lairenjam, C., Huidrom, S., Bandyopadhyay, A., & Bhadra, A. (2016). Assessment of probability distribution of rainfall of North East Region (NER) of India. *J Res Environ Earth Sci*, *2*(9), 12-18.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68-78.

Meena, R. K., Dubey, S., & Basak, S. M. (2019). Use of different probability distribution in rainfall analysis for RRS, Pundiari, Cooch Behar. *Int J Curr Microbiol App Sci*, *8*(10), 1325-1337.

Nwaigwe, C. C., Ogbonna, C. J., & Achem, O. (2023). Statistical Modeling of Rainfall Distribution in Jos, Plateau State, Nigeria. *Asian Journal of Probability and Statistics*, *22*(1), 46-55.

Rag, A., Kumar, M., Kumar, A., Khichar, M. L., Shekhar, C., & Jeet, M. (2025). Deciphering rainfall patterns in Haryana, India: a comprehensive analysis of optimal probability distributions. *Mausam*, *76*(2), 387-402.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Sharma, M. A., & Singh, J. B. (2010). Use of probability distribution in rainfall analysis. *New York Science Journal*, *3*(9), 40-49.

Scott, A. C., & Divakar, (2024). Ecology and agricultural pattern: a geographical study of bundelkhnad region. *The International Journal of Engineering Research*, 11(*1*), 294-303.