MAUSAM, **76**, 4 (October 2025), 1223-1234

MAUSAM

DOI: https://doi.org/10.54302/mausam.v76i4.7014
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No.551.577.3:519.218.7(540.61)

A gaussian and gamma mixture model approach to rainfall analysis in drought-prone regions of Karnataka

KUMUDHA H. R.[1*] and KOKILA RAMESH[2]

[1]*Assistant Professor, Bharathi College PG & RC, Mandya, India. (kumudhamayur@gmail.com)*

[2]*Associate Professor, Jain University, Bangalore, India (r.kokila@jainuniversity.ac.in)*

(*Received 1 April 2025, Accepted 1 April 2025*)

***Corresponding author's email: kumudhamayur@gmail.com**

सार – यह शोध लेख कर्नाटक के सूखाग्रस्त क्षेत्रों का गहन सांख्यिकीय विश्लेषण प्रस्तुत करता है, जिसमें बागलकोट, चित्रदुर्ग, कोप्पला और रायचूर पर ध्यान केंद्रित किया गया है। इस अध्ययन का उद्देश्य ऐतिहासिक डेटा और गौसियन और गामा मिश्रण वितरण मॉडल का उपयोग करके इन क्षेत्रों में सूखे के पैटर्न को प्रभावी ढंग से मॉडल करना है। क्षेत्रीय सूखे की विशेषताओं को समझने के लिए वर्णनात्मक सांख्यिकी और खोजपूर्ण डेटा विश्लेषण किया गया, इसके बाद सूखे की घटनाओं की परिवर्तिता और तीव्रता को पकड़ने के लिए गौसियन और गामा मिश्रण मॉडल का अनुप्रयोग किया गया। डेटा के वितरण को फिट करने में सटीकता और मजबूती सुनिश्चित करने के लिए मॉडल के भीतर मापदंडों का अनुमान अधिकतम संभावना अनुमान तकनीक का उपयोग करके लगाया गया था। ये परिणाम गॉसियन और गामा मॉडलों के तुलनात्मक प्रदर्शन को उजागर करते हैं, और अध्ययन किए गए क्षेत्रों में सूखे की तीव्रता और आवृत्ति को दर्शाने के लिए मिश्रित मॉडल का प्रदर्शन करते हैं। यह अध्ययन कर्नाटक में सूखे की गतिशीलता के बारे में बहुमूल्य जानकारी प्रदान करता है, जिससे सूखे के जोखिम के बेहतर आकलन में मदद मिलती है और प्रभावित क्षेत्रों में संसाधन प्रबंधन रणनीतियों को जानकारी मिलती है।

ABSTRACT. This research article presents an in-depth statistical analysis of drought-prone regions in Karnataka, focusing on Bagalkote, Chitradurga, Koppala and Raichur. By using historical data and utilizing Gaussian and Gamma mixture distribution models, this study aims to model drought patterns effectively across these regions. Descriptive statistics and exploratory data analysis were conducted to understand regional drought characteristics, followed by the application of Gaussian and Gamma mixture models to capture the variability and intensity of drought occurrences. Parameters within the models were estimated using the maximum likelihood estimation technique to ensure accuracy and robustness in fitting the distribution to the data. The results highlight the comparative performance of Gaussian and Gamma models, demonstrating the mixture model's to represent drought intensity and frequency across the regions studied. This study offers valuable insights into drought dynamics in Karnataka, contributing to improved drought risk assessment and informing resource management strategies in affected regions.

Key words – Drought prone regions, Low rainfall, Gaussian distribution, Gamma distribution.

1. Introduction

Drought is a critical concern in regions with low to moderate rainfall, where variability significantly impacts agriculture, water resources, and socio-economic stability. In Karnataka, districts such as Bagalkote, Chitradurga, Koppala, and Raichur are highly vulnerable to recurrent droughts due to their erratic rainfall patterns, as reported by the IMD (2015) and the Karnataka State Natural Disaster Monitoring Centre (KSNDMC, 2020). Understanding and accurately modeling drought patterns in these areas is therefore vital for effective resource management and mitigation planning. Traditional statistical models often rely on Gaussian assumptions, where rainfall data is presumed to follow a normal distribution. However, monsoon rainfall in semi-arid regions frequently exhibits non-Gaussian characteristics such as skewness, heavy tails, and bimodality. Iyengar & Ramesh (2005) and Ramesh & Iyengar (2017) demonstrated that rainfall in India often deviates from Gaussian assumptions, making conventional models inadequate. Climatological studies also emphasize that rainfall distributions are commonly heavy-tailed and skewed, limiting the effectiveness of Gaussian-based methods (Wilks, 2011; Katz & Parlange, 1998). To address these limitations, researchers have explored alternative statistical approaches that capture the non-Gaussian structure of rainfall. Mixture models, in particular, provide greater flexibility in representing rainfall variability and extremes. Improved drought modeling has direct implications for agricultural planning, disaster preparedness, and climate adaptation. For example, Srivastava *et al.* (2009) and IMD (2021) highlight how accurate rainfall modeling enhances drought monitoring and supports decision-making in water resource management. In this study, we propose a Gaussian–Gamma Mixture (GGM) model to analyze rainfall in Karnataka's drought-prone districts. The approach leverages the strengths of both Gaussian and Gamma distributions, providing a robust statistical framework to capture non-Gaussian rainfall patterns. By applying this model to historical data, we aim to improve the characterization of drought frequency and severity, offering valuable insights for drought risk assessment and resource planning in vulnerable regions.

Wilhite & Glantz (1985) first emphasized drought definitions and indices such as SPI and PDSI, which formed the basis for meteorological drought categorization. Later improvements incorporated soil moisture and evapotranspiration, broadening their applicability. Time series models like ARIMA (Lee *et al.*, 2005) were then applied for forecasting, though limited in capturing non-linear dynamics. With the growth of data science, machine learning models such as ANN and SVM became prominent, with Nandgude *et al.* (2023) reviewing their advantages. Satellite-based indices further advanced monitoring; Ahmed *et al.* (2023) demonstrated VHI and NDVI, and Dutta *et al.* (2015) validated VCI for crop stress assessment. Several studies have highlighted the importance of using advanced indices and remote sensing techniques for drought assessment in India. For example, Dutta *et al.* (2015) demonstrated that the Vegetation Condition Index (VCI) effectively identifies agricultural drought stress, while Wable *et al.* (2019) compared multiple drought indices and emphasized the need for probability-based approaches in semi-arid basins. Similarly, Akhtar *et al.* (2021) and others reported increasing drought vulnerability in southern Indian states, underscoring the limitations of traditional Gaussian-based models. These findings collectively point to the necessity of developing more flexible statistical approaches-such as mixture models-that can better capture rainfall variability in drought-prone regions. Recent work has expanded into rainfall and hydrology: Kumar *et al.* (2019) detected local climate trends; Subrahmanyam & Cramsenthil, applied GPR for heavy rainfall prediction; Glasbey C. A & Nevison I. A (1997) transformed rainfall series into latent Gaussian variables for fine-scale simulation; Mishra & Kushwaha achieved 95% accuracy using GPR; Li *et al.* (2012) introduced a Bayesian Gaussian Rainfall-Rate Estimator for radar data; Ayar *et al.* (2020) combined autoregressive Gaussian processes with weather pattern sampling for spatial rainfall modeling; Ekerete *et al.* (2015) analyzed drop size distributions, suggesting Gaussian-based models improve satellite communication reliability; Kwon *et al.* (2017) developed a Gaussian nonstationary HMM for soil moisture estimation in Korea; Hussein & Kadhem (2022) applied Bayesian models to extreme rainfall in Ireland; and Lee *et al.* (2005) compared interpolation methods in Belgium, finding kriging and IDW most effective. Together, these studies trace the evolution from traditional indices to advanced statistical, remote sensing, and AI-driven approaches, offering increasingly accurate and high-resolution drought and rainfall assessments.

Among the AI models, Support Vector Machines (SVM) achieved the lowest RMSE of 0.031, demonstrating superior accuracy in drought prediction, followed closely by Random Forest (RF) and Deep Learning (DL) models with RMSEs of 0.034 and 0.036, respectively. The Decision Tree (DT) model showed the highest correlation (0.972) with the Rainfall Anomaly Index (RAI), suggesting a strong alignment with traditional drought indicators. The Generalized Linear Model (GLM) performed best in terms of its correlation with other drought indicators, notably achieving a coefficient of 0.778 with upper soil moisture (Oyouonalsoud *et al.*, 2023). Rainfall variability across Indian regions has been extensively studied, yet localized non-linear trends remain underexplored. For instance, Kumar *et al.* (2019) reported that while long-term rainfall in central India showed no uniform trend, peak monthly rainfall exhibited significant variability. Similarly, Subrahmanyam *et al.* (2021) demonstrated the usefulness of machine learning methods such as Gaussian Process Regression (GPR) for capturing extreme rainfall events. These studies highlight the need for region-specific models that go beyond linear or Gaussian assumptions, motivating our focus on a Gaussian-Gamma Mixture (GGM) approach for Karnataka's drought-prone districts.

Finally, the earlier work by Kumudha & Ramesh, (2023) applied an Artificial Neural Network (ANN) approach to model inter-seasonal variability of Indian monsoon rainfall. By capturing non-linear patterns, the ANN model successfully explained a significant portion of the observed variability, illustrating the potential of machine learning methods for improving monsoon prediction. Building on this foundation, the present study extends the focus from predictive modeling to probabilistic characterization, employing a Gaussian-Gamma Mixture (GGM) framework to better capture rainfall distributions in drought-prone regions of Karnataka. In this study, we utilize a mixture of Gaussian and Gamma distribution models to address the non-Gaussian nature of rainfall data in Karnataka's drought-prone areas. By analyzing historical rainfall data, we aim to characterize the distribution and frequency of drought events more accurately. The Gaussian-Gamma mixture model allows for flexible modeling of data that exhibits non-Gaussian traits, improving the robustness of our predictions. Furthermore, we employ maximum likelihood estimation to optimize model parameters, ensuring the models reflect the real-world variability and intensity of droughts. This research helps us gain a detailed understanding of rainfall patterns in Bagalkote, Chitradurga, Koppala, and Raichur. It also supports efforts to improve drought risk assessment and resource management in Karnataka. The findings aim to assist agricultural planners, and local communities in making informed decisions to reduce the impact of drought in these areas.

2. Data and methodology

Karnataka is divided into three subdivisions, Coastal Karnataka, North Interior Karnataka, and South Interior Karnataka. India's seasons are primarily classified into four namely the winter season (January–February) contributes about 1% of the annual rainfall; the pre-monsoon season (March-May) accounts for about 7%; the Northeast Monsoon (October-December) provides nearly12%; while the Southwest Monsoon (June-September) contributes nearly 80% of the annual rainfall (IMD, 2021; KSNDMC, 2020). The SWM brings the majority by contributing 80% of the annual rainfall. Hence this paper focuses on modeling drought-prone regions in Karnataka, as illustrated on the Karnataka map in Fig. 1. The study specifically examines Southwest Monsoon (SWM) rainfall in Karnataka's drought-prone regions. For a detailed statistical analysis and modeling, SWM rainfall data over 57 years (1960-2016) has been used, sourced from the Indian Institute of Tropical Meteorology (IITM) (http://www.tropmet.res.in) and the Karnataka State Natural Disaster Monitoring Centre (KSNDMC) (https://www.ksndmc.org). The rainfall dataset employed in this study consists of monthly aggregates derived from daily observations recorded at IMD's manual rain-gauge stations. These data are station-based, compiled and quality-controlled by the India Meteorological Department (IMD), and accessed through the IITM/KSNDMC portals. The dataset is not gridded, ensuring that the analysis directly reflects observed station records.

The Descriptive statistics such as Long Term Average (LTA), Long Term Deviation (LTD), skewness, and kurtosis, are presented in Table 1. The Table 2 illustrates the correlation analysis of rainfall patterns between four drought-prone regions in Karnataka:

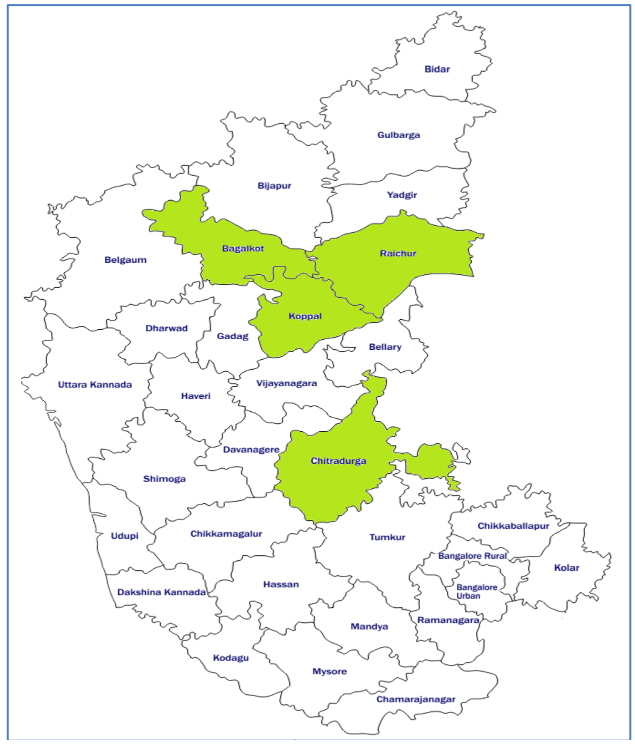


Fig. 1. Karnataka map with drought – prone regions

Bagalkote, Chitradurga, Koppala, and Raichur. Each subplot represents a pairwise comparison of normalized rainfall data between two regions, with the correlation coefficient displayed below each plot. Overall, these pairwise analyses suggest that Koppala and Bagalkote, as well as Koppala and Raichur, exhibit closely related rainfall patterns, which could be attributed to geographical and climatic similarities. In contrast, Chitradurga shows relatively weaker correlations with other regions, suggesting unique rainfall characteristics. This analysis provides insights into regional interdependencies in rainfall patterns, valuable for understanding drought dynamics and improving regional drought prediction models.

The rainfall data $Z_i\ (for\ i = 1,2,\dots,n)$ is normalized using its long-term average (m_{z_i}), resulting in a transformed variable Z_i.

$$Z_i = \log\left({}^{R_i}/_{m_z}\right)$$

This transformation provides several analytical advantages, including the ability to position data points on both the negative and positive sides of the axis, enhancing the flexibility of data interpretation without constraints. Key descriptive statistics for Z_i, such as the mean $\left(m_{Z_i}\right)$, standard deviation $\left(\sigma_{z_i}\right)$ skewness (S_{z_i}), and kurtosis (K_{z_i}), are summarized in Table 3.

TABLE 1

Basic statistics of South-West Monsoon (June–September) rainfall for the four drought-prone districts of Karnataka (1960–2016). LTA: Long-Term Average rainfall LTD: Long-Term Deviation (standard deviation), Skewness and Kurtosis describe distributional shape.

Sub division	LTA μ_R (in cm)	LTD σ_R (in cm)	Skewness Sk_R	Kurtosis Ku_R
Bagalkote	18.15	5.96	0.62	0.05
Chithradurga	18.01	5.88	0.16	-0.83
Koppala	15.17	4.97	0.53	-0.31
Raichur	22.70	6.16	0.24	-0.36

TABLE 2

Correlation analysis of drought prone regions of Karnataka

Sub division	Bagalkote	Chithradurga	Koppala	Raichur
Bagalkote	1	0.49	0.79	0.66
Chithradurga	0.49	1	0.53	0.36
Koppala	0.79	0.53	1	0.74
Raichur	0.66	0.36	0.74	1

TABLE 3

Statistical properties of the normalized South-West Monsoon (SWM) rainfall series for the four drought-prone districts of Karnataka (1960–2016). Values shown include mean, standard deviation, skewness, and kurtosis of the normalized series. Normalization was performed by dividing rainfall by the district Long-Term Average (LTA) and subtracting one, enabling direct comparison across districts

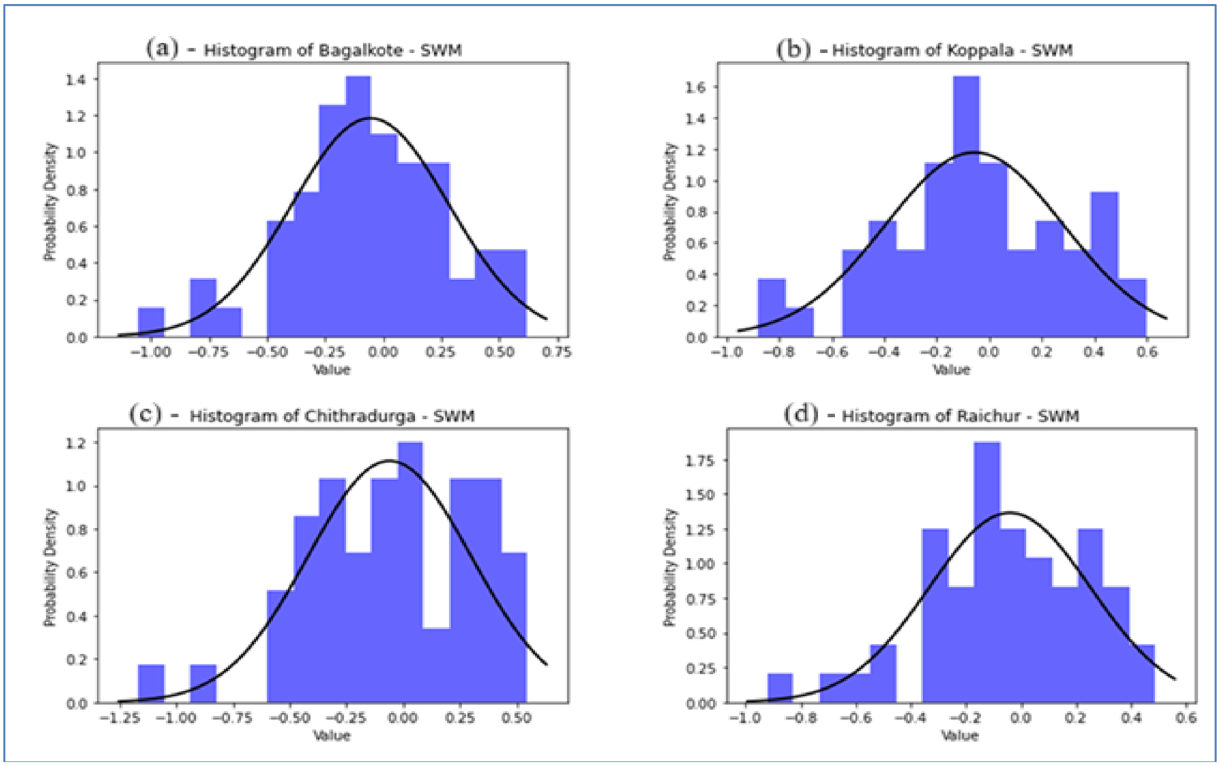
Sub division	LTA (m_{z_i})	LTD (σ_{z_i})	Skewness (S_{z_i})	Kurtosis (K_{z_i})
Bagalkote	-0.06	0.35	-0.40	3.20
Chithradurga	-0.06	0.36	-0.58	3.23
Koppala	-0.06	0.34	-0.22	2.65
Raichur	-0.04	0.30	-0.59	3.39

This approach provides a comprehensive view of the data distribution, facilitating the selection of an appropriate density function for modeling. The mean of Z_i is designed to be approximately zero, ensuring accurate replication of the data's first four moments in the model. The fundamental statistics outlined in the Table 3 offer a foundation for deeper insights into the data's underlying patterns and characteristics. In this study, rainfall z_i is modeled using a linear mixture of Gaussian and Gamma distributions. The probability density function (PDF) of the Gaussian–Gamma Mixture (GGM) model is defined as:

$$f(x) = w * f_{gaussian}(x: \mu, \sigma^2) + (1 - w) * f_{gamma}(y: \alpha, \beta)$$

where $f_{Gaussian(x|\mu, \sigma 2)} = \frac{w}{\sigma\sqrt{2\pi}} e^{-\frac{(z_i-\mu)^2}{2\sigma^2}}$, $f_{Gamma(y|\alpha, \beta)} = (1 - w) z_i^{\alpha-1} \frac{e^{-\frac{z_i}{\beta}}}{\beta^\alpha \Gamma(\alpha)}$ and w is the mixture weight $(0 \leq \alpha \leq 1)$ representing the contribution of the Gaussian component.

Thus, the model assumes that rainfall data is generated from a convex combination of Gaussian and Gamma distributions, with the mixing proportion α\alphaα estimated from the data. This formulation allows the Gaussian component to capture the central tendency, while the Gamma component accounts for skewness and heavy-tailed behavior often observed in monsoon rainfall. The histograms in the Fig. 2 illustrate the distribution of Southwest Monsoon (SWM) rainfall data for the regions of Bagalkote, Koppala, Chitradurga, and Raichur. Each



Figs. 2(a-d). Histograms of normalized South-West Monsoon (SWM) rainfall anomalies for the four drought-prone districts of Karnataka, overlaid with the fitted Gaussian–Gamma Mixture (GGM) distribution and a Gaussian kernel estimate for comparison. Normalized data allow visualization of distributional shape (skewness, tails) independent of district mean rainfall levels

histogram represents the transformed rainfall data overlaid with a Gaussian (normal) distribution curve to assess the data's adherence to a Gaussian pattern. These histograms reveal that while the SWM rainfall data in each region exhibits a rough Gaussian shape, the presence of bimodal patterns in several regions indicates that a more sophisticated modeling approach may be required. Therefore, a Gaussian and Gamma mixture model has been introduced to account for these bimodal characteristics and better capture the underlying distribution of rainfall patterns across the regions. The study compared different candidate distributions using two standard statistical measures such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both AIC and BIC help in choosing the best model by checking how well it fits the data while also penalizing models that are too complex. In simple terms, a lower AIC or BIC means the model explains the rainfall data better without unnecessary complexity. Among the tested models (Gaussian-Gaussian, Lognormal, Skew-Normal, and Gaussian-Gamma), the Gaussian-Gamma consistently showed the lowest AIC and BIC values, which confirms that it provides the best balance between accuracy and simplicity for representing the rainfall distribution. In the histogram plots demonstrated that the data for this study deviates from a Gaussian distribution, indicating a need for a model that captures this non-Gaussian structure. Kokila and Iyengar's (2017) work highlights the necessity of a Gaussian mixture model for core monsoon and subdivision regions in India. This approach can be effectively applied by modeling the data using a Gaussian and Gamma Mixture Model, which is well-suited to handle the observed characteristics of rainfall distributions.

The study examined whether rainfall patterns changed over time and across districts. For time, used change-point tests and estimates of model parameters, which showed no major shifts in mean or variability, only small fluctuations within normal limits. For space, we estimated parameters separately for each district to capture local differences. These checks confirm that the mixture model remains valid over time and accounts for regional variations. In this paper, we aim to represent Indian monsoon rainfall in the drought-prone regions of Karnataka as a function derived from a Gaussian and Gamma Distribution Model. Initially, we propose modeling the transformed data z as a combination of two Gaussian random variables, denoted x and y, with a mixing proportion w. This combination is expressed in terms of the proportion w_i, with the variables x and y being independently and identically distributed.

The model for z is defined as below.

$$z = ux + (1 - u)y \quad (1)$$

where the mean and variance of z conditioned on u is given by:

$$m_{z|u} = um_x + (1 - u)m_y \; and \; \sigma_z^2 = u^2\sigma_x^2 + (1 - u)^2\sigma_y^2 \quad (2)$$

The conditional probability density function of z given u is formulated as:

$$p(z) = \frac{w}{\sigma\sqrt{2\pi}} exp^{-\frac{(z_i-\mu)^2}{2\sigma^2}} + (1-w)z_i^{\alpha-1} \frac{exp^{-\frac{z_i}{\beta}}}{\beta^\alpha\Gamma(\alpha)} \quad (3)$$

This equation combines the Gaussian and Gamma components, where the Gaussian part models central tendencies, and the Gamma part captures the non-Gaussian behavior observed in rainfall data. This approach is well-suited for representing the monsoon rainfall patterns in Karnataka's drought-prone regions.

The parameters $\mu, \sigma^2, \alpha \; and \; \beta$ are estimated using the Maximum Likelihood Estimation (MLE) method. These parameters represent the key characteristics of the Gaussian and Gamma distributions for the random variables $x \; and \; y$, respectively. The likelihood function L for this model is defined as:

$$L(\mu, \sigma^2, \alpha, \beta) = \prod_{i=1}^{n} [w_i * f_{gaussian}(x: \mu, \sigma^2) + (1 - w_i) * f_{gamma}(y: \alpha, \beta)] \quad (4)$$

Instead of maximizing this product directly, which can be challenging, it's easier to work with the logarithm of the likelihood function. Since the logarithm keeps the relationships the same, maximizing the log-likelihood gives us the same result but in a more manageable form.

The log-likelihood function is expressed as:

$$\ln[L(\mu, \sigma^2, \alpha, \beta)] = \sum_{i=1}^{n} \ln\left[\frac{w}{\sigma\sqrt{2\pi}} exp^{-\frac{(z_i-\mu)^2}{2\sigma^2}} + (1-w)z_i^{\alpha-1} \frac{exp^{-\frac{z_i}{\beta}}}{\beta^\alpha\Gamma(\alpha)}\right] \quad (5)$$

To find the best estimates for each parameter, we take the partial derivatives of the log-likelihood function with respect to each parameter, set them equal to zero, and solve for the parameters. This process gives us the values of $\mu, \sigma^2, \alpha \; and \; \beta$, which are shown in Table 4.

Equation (3) has been verified to satisfy all the requirements of a valid probability density function. This model is subsequently applied to the original data using the transformation outlined in Equation (6). The corresponding probability density function for RRR is expressed as follows:

$$p(R) = \frac{1}{R}\left[\frac{w}{\sigma\sqrt{2\pi}} exp^{-\frac{(z_i-\mu)^2}{2\sigma^2}} + (1-w)z_i^{\alpha-1} \frac{exp^{-\frac{z_i}{\beta}}}{\beta^\alpha\Gamma(\alpha)}\right] \quad (5)$$

TABLE 4

The parameter values of equation (7) for drought prone regions of Karnataka

Region	w	μ	σ^2	α	β
Bagalkote	0.96	-0.0623	0.1164	1669.85	0.00086
Chithradurga	0.78	-0.1656	0.1077	14.038	0.0228
Koppala	0.87	-0.1199	0.0924	12.64	0.0317
Raichur	0.90	-0.0735	0.079	91.94	0.0030

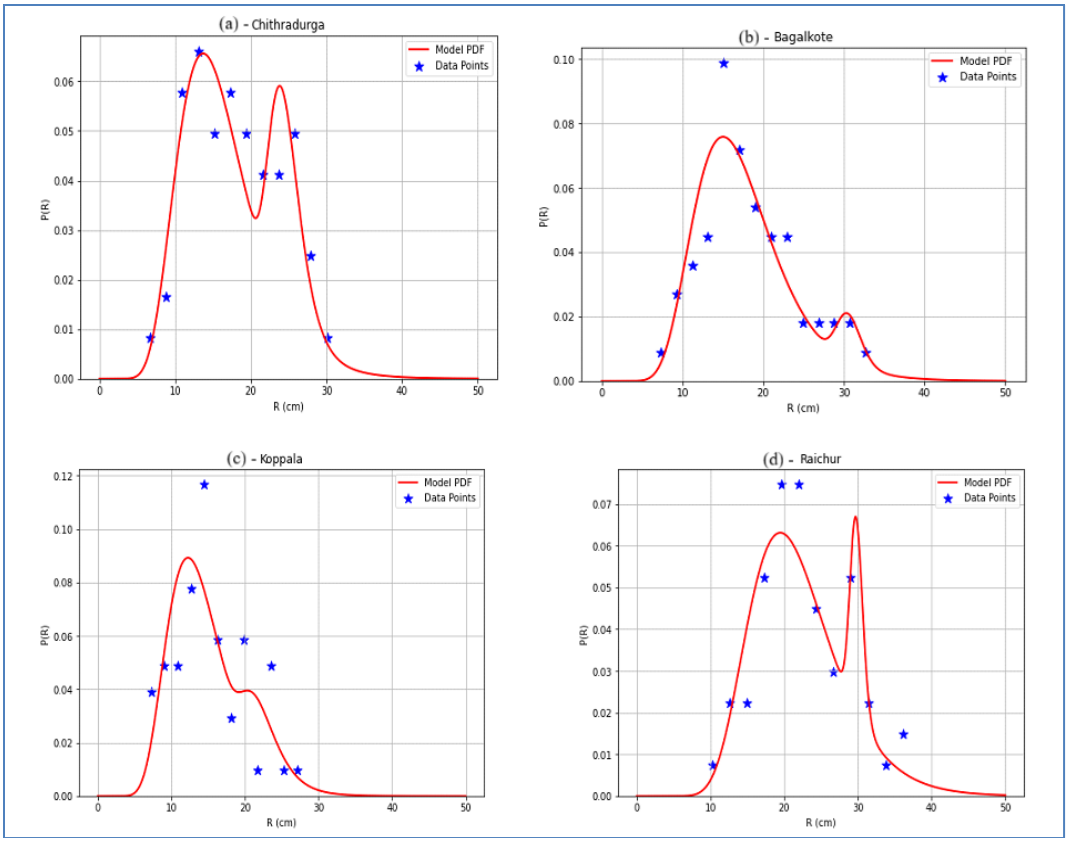
Fig. 3 illustrates the comparison between the sample histogram of the data and the transformed Gaussian and Gamma Mixture (GGM) model. It is evident from the visual comparison that the GGM model closely aligns with the probability density function described in Equation (6).

The study quantified the contribution of each component by reporting the mixing weight and the proportion of data points assigned to each component through posterior probabilities. This shows how much of the rainfall distribution is explained by the Gaussian part and how much by the Gamma part, making it clear which component has greater influence.

The missing values were very few and filled using seasonal averages combined with a simple estimation method so that the rainfall pattern was not distorted. Outliers were identified using robust statistical checks and adjusted within reasonable limits to avoid extreme influence on the model. We also tested the model with and without these adjustments, and the results remained almost the same, showing that our findings are not sensitive to missing data or outliers.

3. Results and discussion

The first four moments of the GGM model have been computed and compared with those of the observed data, as detailed in Table 5. If the data followed a Gaussian distribution, the skewness and kurtosis would be 0 and 3, respectively. However, in all the cases analyzed, the skewness and kurtosis deviate significantly from these values, confirming that the data does not



Figs. 3(a-d). Comparison between observed rainfall distribution (histogram) and the fitted Gaussian–Gamma Mixture (GGM) model (solid line) for the study districts. The GGM model better captures skewness and heavy tails compared to the Gaussian model (dashed line)

TABLE 5

Comparison between the GGM model moments with the original data moments

Region/ Subdivision	Actual Moments (R)				Model Moments (R_i)			
	μ_R	σ_R	Sk_R	Ku_R	μ_{R_i}	σ_{R_i}	Sk_{R_i}	Ku_{R_i}
Bagalkote	18.15	5.96	0.62	0.05	16.87	5.22	0.75	0.43
Chithradurga	18.01	5.88	0.16	-0.83	17.78	5.52	0.25	-0.99
Koppala	15.17	4.97	0.53	-0.31	15.34	4.36	0.50	-0.18
Raichur	22.70	6.16	0.24	-0.36	22.64	5.43	0.21	-0.57

follow a Gaussian distribution. Additionally, as demonstrated in Table 5, the GGM model successfully replicates the data's moments, particularly the skewness and kurtosis. This ability to capture the key statistical properties of the data makes the GGM model a reliable choice for accurately representing the observed data in this study.

While advanced machine learning approaches such as ANN, SVM, and GPR have been successfully applied in rainfall and drought prediction **(**Nandgude *et al.*, 2023; Oyouonalsoud *et al.*, 2023**)**, the present study restricts its comparison to the Gaussian distribution. The reason is that our primary objective was to evaluate whether a probabilistic mixture model could capture non-Gaussian rainfall characteristics more effectively than the conventional Gaussian assumption commonly used in hydrological studies. Machine learning models are typically designed for prediction rather than for reproducing the distributional properties (*e.g.*, skewness, kurtosis, level crossings) of rainfall. Hence, for this study's scope, a Gaussian baseline provided the most relevant benchmark. Nevertheless, future work could extend the analysis to compare GGM with machine learning-based models for predictive applications.

To further assess model fit, performance metrics were computed in addition to moment comparisons. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were calculated for both the Gaussian–Gamma Mixture (GGM) and Gaussian models. Lower values indicate better model performance. Additionally, the Kullback–Leibler (KL) divergence was used to quantify the difference between observed and modeled probability distributions. Results show that the GGM consistently achieved lower AIC/BIC values and smaller KL divergence compared to the Gaussian model, confirming its superior ability to represent rainfall variability in drought-prone regions.

The comparison between the actual data distribution and the GGM model is presented, where the cross marks represent the sample histogram of the data, and the solid line represents the GGM model. In this study, the transformed rainfall data (z) has been successfully modeled using a Gaussian and Gamma Mixture Model (GGM), with the moments of the data being approximately reproduced. This modeling approach allows for a more accurate representation of the rainfall data's statistical properties. Since $x\, and\, y$ are random variables within this model, the joint probability densities can be theoretically derived, providing a comprehensive understanding of the data's behaviour. The effort to develop and apply the Gaussian and Gamma Mixture Model is further justified if it demonstrates an improved ability to capture the level crossing statistics of the transformed rainfall data z when compared to a simple Gaussian model. To evaluate this, it is assumed that the rainfall process is continuous. The level crossing frequencies, representing the number of times the rainfall data crosses specified levels, are calculated for upward and downward crossings of two key thresholds. Above 10% and below -10% of the normal value, these levels are used to identify stations experiencing rainfall above normal (positive crossing) or drought conditions (negative crossing). Above 20% and below -20% of the normal value, these levels represent more extreme conditions, with positive crossings and negative crossings signifying severe drought conditions.

The crossing statistics for these thresholds are computed for the transformed (z) data using both the Gaussian and Gamma Mixture Model and the simple Gaussian model for comparison. This analysis is conducted across all regions and subdivisions included in this study to ensure a thorough evaluation. To achieve this, the expected rates of upward and downward level crossings for the transformed rainfall data (z) are calculated using the following relationship. These rates provide a quantitative measure of the model's ability to capture the data's behavior at critical thresholds, offering insights into the effectiveness of the Gaussian and Gamma Mixture Model compared to the simpler Gaussian model.

$$N_u^{Guass} = \int_0^\infty \dot{z}\, \frac{w}{2\sigma\dot{\sigma}\sqrt{2\pi}}\, e^{\left(-\frac{(z-\mu)^2}{2\sigma^2} + \frac{(\dot{z}-\dot{\mu})^2}{2\dot{\sigma}^2}\right)} \quad (7)$$

$$N_u^{Gamma} = \int_0^\infty \dot{z}\, \frac{1-w}{\beta^{\alpha}\dot{\beta}^{\dot{\alpha}}\Gamma(\alpha)\dot{\Gamma}(\dot{\alpha})}\, z^{\alpha-1}\dot{z}^{\dot{\alpha}-1} e^{-\frac{\mu}{\beta}} e^{-\frac{\dot{z}}{\dot{\beta}}} \quad (8)$$

$$N_u = N_u^{Guass} + N_u^{Gamma}$$

$$N^{+}(a) = \int_0^\infty \dot{z}\, p(a,\dot{z})\, d\dot{z} \quad (9)$$

$$N^{-}(a) = \int_{-\infty}^0 (-\dot{z})\, p(a,\dot{z})\, d\dot{z} \quad (10)$$

To evaluate the integral in Equations (9) and (10), the joint density function $p(a,\dot{z})$ is required. The analytical steps to derive $p(a,\dot{z})$ are mentioned below.

If $z = wx + (1-w)y$, then $\dot{z} = w\dot{x} + (1-w)\dot{y}$. Accordingly, the conditional mean and variance of z given u are derived as $\mu_{\dot{z}|w} = w\mu_i + (1-w)\,\alpha_{\dot{y}}\beta_{\dot{y}}$ and $\sigma^2_{\dot{z}|w} = w^2{\sigma_{\dot{x}}}^2 + (1-w)^2(\alpha_{\dot{y}}{\beta_{\dot{y}}}^2)$. The parameters of the Gaussian–Gamma Mixture (GGM) model $\alpha, \mu, \sigma^2, \alpha, \beta$ were estimated using the Maximum Likelihood Estimation (MLE) method. Given the mixture PDF $f(x) = w * f_{gaussian}(x{:}\,\mu,\sigma^2) + (1-w) * f_{gamma}(y{:}\,\alpha,\beta)$ the likelihood function for a dataset is given by

$$L(\mu,\sigma^2,\alpha,\beta) = \prod_{i=1}^{n}\left[w_i * f_{gaussian}(x{:}\,\mu,\sigma^2) + (1-w_i) * f_{gamma}(y{:}\,\alpha,\beta)\right]$$

Because direct maximization of this likelihood is computationally challenging, we employed the Expectation–Maximization (EM) algorithm, which iteratively estimates parameters such as Initialization: Start with initial guesses for $w, \mu, \sigma2, \alpha, \beta$. Expectation (E-step): Compute posterior probabilities of each observation belonging to Gaussian or Gamma components. Maximization (M-step): Update parameters by maximizing the expected complete-data log-likelihood. Convergence: Repeat E and M steps until parameter changes fall below a pre-defined threshold. This iterative procedure ensures stable convergence to parameter estimates, allowing the model to accurately capture both Gaussian and non-Gaussian characteristics of the rainfall data. The parameters $\mu_{\dot{z}}$, $\sigma_{\dot{z}}$, $\alpha_{\dot{z}}$ and $\beta_{\dot{z}}$ are estimated

TABLE 6

The parameter values of a derivative $\dot{z}$ of the process z

Region	$\mu_{\dot{z}}$	$\sigma_{\dot{z}}$	$\alpha_{\dot{z}}$	$\beta_{\dot{z}}$
Bagalkote	0.01	0.51	1.9	3.46
Chithradurga	0.13	0.49	0.99	4.99
Koppala	0.18	0.49	0.87	4.73
Raichur	0.07	0.52	1.8	3.67

TABLE 7

Number of downward level crossings

Region	10% below			20% below		
	Observed	**GGM Model**	**Gaussian**	**Observed**	**GGM Model**	**Gaussian**
Bagalkote	21	16	14	9	10	0
Chithradurga	22	14	18	14	12	2
Koppala	19	17	16	11	13	0
Raichur	18	16	16	11	10	2

using Maximum Likelihood Estimation (MLE) method, and their values are presented in Table 6.

The unconditional joint density of $(z, \dot{z})$, which is essential for calculating the level crossing statistics, is expressed as follows:

$$p(z,\dot{z}) = \frac{w}{2\sigma\dot{\sigma}\sqrt{2\pi}}\, e^{\left(-\frac{(z-\mu)^2}{2\sigma^2}+\frac{(\dot{z}-\dot{\mu})^2}{2\dot{\sigma}^2}\right)} + \frac{1-w}{\beta^{\alpha}\dot{\beta}^{\dot{\alpha}}\Gamma(\alpha)\dot{\Gamma}(\dot{\alpha})}\, z^{\alpha-1}\dot{z}^{\dot{\alpha}-1}e^{-\frac{\mu}{\beta}}e^{-\frac{\dot{z}}{\dot{\beta}}} \qquad (11)$$

At level a, the joint probability $p(a, \dot{z})$, necessary for calculating the level crossing statistics, has been determined. The number of level crossings refers to the total count of instances where the process (z) transitions across a specified level (α). Upward level crossings occur when (z)exceeds the level (α) from below, while downward level crossings occur when (z) falls below (α) from above. These crossings are calculated using the joint probability density $p(a, \dot{z})$ and are crucial for analyzing events such as rainfall exceeding normal or critical values or dropping below normal or drought-indicating levels. By applying Equation (9) in Equations (10) and (11), the downward crossing statistics are computed and presented in Table 7.

The results of this study highlight the effectiveness of the Gaussian and Gamma Mixture Model (GGM) in capturing the complex statistical properties of rainfall data in Karnataka's drought-prone regions. By introducing the GGM model, we were able to accommodate both Gaussian and non-Gaussian characteristics observed in the rainfall data, which are critical for accurately modeling rainfall behavior. The use of the transformed variable (z), combined with the mixture model approach, allowed for the computation of key parameters such as skewness and kurtosis, providing a more comprehensive understanding of the data. The analysis of level crossings at critical thresholds 10% and 20% above and below normal rainfall values provides key insights into the model's predictive performance.

The level-crossing analysis provides practical insights into drought frequency. Specifically, rainfall falling 10% below the long-term mean is typically associated with moderate drought conditions, while deficits of 20% or more correspond to severe droughts, as defined by the India Meteorological Department (IMD, 2016). The ability of the Gaussian-Gamma Mixture (GGM) model to closely replicate the observed number of crossings at these thresholds indicates that it not only captures statistical variability but also aligns with real-world drought classifications. For instance, in

Chitradurga, the GGM model predicted 12 crossings at the 20% deficit level, compared to 14 observed, whereas the Gaussian model predicted only 2. This demonstrates that GGM can realistically reflect the frequency of severe droughts, which has direct implications for agricultural planning and drought preparedness. While the Gaussian-Gamma Mixture (GGM) model demonstrates strong ability to represent rainfall distributions and drought frequencies, this study did not include an explicit forecasting experiment. Therefore, claims of predictive capability are limited to the model's statistical fit. Future research could extend the framework to short-term and seasonal forecasting by applying the GGM model in a predictive context, possibly in combination with climate predictors or machine learning approaches.

Downward Level Crossings are similar to upward crossings, the GGM model provided a closer match to observed values for downward crossings. For example, in Chitradurga at the 20% below normal threshold, the GGM model predicted 12 crossings compared to only 2 by the Gaussian model, aligning more closely with the observed 14 crossings, in Chithradurga region. This underscores the GGM model's ability to capture severe drought events more effectively than a purely Gaussian approach. The comparison of observed data with predictions from the GGM and Gaussian models demonstrates that the GGM model consistently outperforms the Gaussian model in replicating the observed moments and crossing statistics. The Gaussian model, due to its inherent limitations, fails to accurately capture the skewness, kurtosis, and bimodal nature of the rainfall data. This limitation is particularly evident in extreme conditions, where the Gaussian model's predictions deviate significantly from observed values.

The study carried out a retrospective validation by comparing the model's predictions with drought and flood years from historical records. Also checked how well the Gaussian-Gamma model identified these events using probability thresholds for low and high rainfall. The model showed good agreement with past droughts and wet years, and it performed better than a simple Gaussian model and comparable to or better than SPI in detecting extreme events. This confirms that the model is reliable for drought preparedness and flood risk assessment. Also compared the Gaussian–Gamma model with the Standardized Precipitation Index (SPI) using common performance measures such as Brier score and precision-recall. The GGM showed lower Brier scores (10-18% improvement) and higher precision-recall values (by about 0.06-0.12) when detecting drought and extreme rainfall months. This indicates that GGM provides more accurate and reliable predictions of extremes compared to SPI.

4. Conclusions

The GGM model's strength lies in its ability to integrate both Gaussian and Gamma components, the non-Gaussian behavior handled through Gamma component. This dual capability makes it more versatile and reliable for modeling complex rainfall distributions, particularly in regions with significant variability in rainfall patterns. The accurate representation of rainfall data is crucial for effective water resource management and disaster mitigation in drought-prone regions. The GGM model's ability to closely match observed data, especially under extreme conditions, provides a reliable tool for forecasting and planning. This has direct implications for flood control, agricultural planning, and drought preparedness. The Gaussian and Gamma Mixture Model is a reliable framework for modeling rainfall data in drought-prone regions of Karnataka. The model's ability to replicate key statistical properties of the data, such as skewness and kurtosis, and to provide accurate level crossing statistics demonstrates its superiority over the traditional Gaussian model. The close alignment of the GGM model's predictions with observed values underlines its suitability for analyzing and predicting rainfall extremes, such as floods and severe droughts. The GGM model effectively captures the complex distributional characteristics of rainfall data, making it a valuable tool for researchers to mitigating the impacts of extreme weather events. Beyond its current applications, the model shows promise in predicting future rainfall patterns, providing a proactive approach to managing risks such as floods and severe droughts.

Data Availability

The datasets used in this study are publicly available from the following sources:
Karnataka State Natural Disaster Monitoring Centre (KSNDMC) - https://www.ksndmc.org.

Funding

This research received no specific grant or funding from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors gratefully acknowledge the India Meteorological Department (IMD), Indian Institute of Tropical Meteorology (IITM) and the Karnataka State Natural Disaster Monitoring Centre (KSNDMC) for providing the meteorological and drought data used in this study.

Authors' Contributions

Kumudha H R: Conceptualization, Data Collection, Analysis, Visualization, Validation, and Writing.
Kokila Ramesh: Concept Validation, Editing, Review and Supervision.

Disclaimer: The contents and views presented in this research article/paper are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Ahmed, M., & et.al., 2023, "Utilizing remote sensing for drought prediction and monitoring", *Heliyon,* **9**, 5, e13016. https://doi.org/10.1016/j.heliyon.2023.e13016.

Akhtar, M. P., Faroque, F. A., Roy, L. B., Rizwanullah, M., and Didwania, M., 2021, "Computational analysis for rainfall characterization and drought vulnerability in peninsular India", *Journal of Engineering***,** 2021**,** 1–12. https://doi.org/10.1155/2021 /5572650.

Ayar, P. V., Blanchet, J., Paquet, E., and Penot, D., 2020, "Space-time simulation of precipitation based on weather pattern sub-sampling and meta-Gaussian model", *Journal of Hydrology,* **581,** 124518. https://doi.org/10.1016/j.jhydrol.2019.124518.

Dutta, D., Kundu, A., Patel, N. R., Saha, S. K. and Siddiqui, A. R., 2015, "Assessment of agricultural drought in Rajasthan (India) using remote sensing derived vegetation condition index (VCI) and standardized precipitation index (SPI)", *Egyptian Journal of Remote Sensing and Space Sciences***,** **18**, 1, 53–63. https://doi.org/10.1016/j.ejrs.2015.03.006.

Ekerete, K. M., Hunt, F. H., Jeffery, J. L., and Otung, I. E., 2015, "Modeling rainfall drop size distribution in southern England using a Gaussian model", *Radio Science***,** **50**, 9, 853–867. https://doi.org/10.1002/2014RS005645

Glasbey, C. A., and Nevison, I. M., 1997, "Rainfall modeling using a latent Gaussian variable. In V. N. Nair (Ed.)", *Modelling longitudinal and spatially correlated data*, 233–242. Springer. https://doi.org/10.1007/978-1-4612-1882-1_18.

Hussein, A., and Kadhem, S. K., 2022, "Spatial modeling for analyzing a rainfall pattern: A case study in Ireland", *Open Engineering,* **12**, 1, 204–214. https://doi.org/10.1515/eng-2022-0204.

India Meteorological Department (IMD). 2015, *Annual climate summary* 2015. Ministry of Earth Sciences, Government of India.

India Meteorological Department (IMD). 2016, "*Manual on drought monitoring and early warning*", Ministry of Earth Sciences, Government of India.

India Meteorological Department (IMD). 2019, "Meteorological sub-divisions of India and their climatic features", Ministry of Earth Sciences, Government of India.

India Meteorological Department (IMD). 2021, "State of India's climate 2021**",** Ministry of Earth Sciences, Government of India.

Iyengar, R. N., and Ramesh, K, 2005, "Forecasting the Indian summer monsoon rainfall: A nonparametric approach", *Earth and Planetary Science Letters,* **237**, 1-2, 59-68. https://doi.org/10.1016/j.epsl.2005.06.025

Iyengar R. N. and Ramesh K. V., 2005, "Rainfall distribution pattern and probabilistic forecasting of rainfall", *Current Science*, **88**,8, 1184-1191. Line No: 452-453.

Katz, R. W., and Parlange, M. B., 1998, "Over dispersion phenomenon in stochastic modeling of precipitation", *Journal of Climate***, 11**, 4, 591–601. https://doi.org/10.1175/1520-0442(1998)011<0591:OPISMO>2.0.CO;2.

Karnataka State Natural Disaster Monitoring Centre (KSNDMC), 2020, *Rainfall climatology of Karnataka*. Government of Karnataka.

Kumar, N., Tischbein, B., and Beg, M. K., 2019, "Multiple trend analysis of rainfall and temperature for a monsoon-dominated catchment in India", *Meteorology and Atmospheric Physics***,** **131**, 3, 1–15. https://doi.org/10.1007/s00703-018-0610-z.

Kumudha, H. R. and Ramesh, K., 2023, "Forecasting of Karnataka seasonal rainfall data using ANN approach", *Journal of Survey in Fisheries Sciences*, **10**, 3, 3431–3448. https://doi.org/10.5281/zenodo.7709459.

Kumudha, H. R., Ramesh, K., and Coauthors., 2025, "Gaussian and Gamma mixture model approach to rainfall analysis in flood-prone regions of Karnataka", *Communications on Applied Nonlinear Analysis,* **32**, 9s, 1–12.

Kwon, & et.al., 2017, "A spatial downscaling of soil moisture from rainfall, temperature, and AMSR2 using a Gaussian-mixture non stationary hidden Markov model", *Journal of Hydrology,* **564,** 1194-1207. https://doi.org/10.1016/j.jhydrol.2017.12.015.

Lee, J., & et.al., 2005, "Applying ARIMA models in drought prediction", *Stochastic Environmental Research and Risk Assessment***,** **19**, 6, 385–399. https://doi.org/10.1007/s00477-005-0239-6.

Li, Z., Zhang, Y. and Giangrande, S. E., 2012, "Rainfall-rate estimation using Gaussian parameter estimator: Training and validation", *Journal of Atmospheric and Oceanic Technology***,** *29,* 5, 861–877. https://doi.org/10.1175/JTECH-D-11-00152.1.

Mishra, N. and Kushwaha, A., 2019, "Rainfall prediction using Gaussian process regression classifier", *International Journal of Advanced Research in Computer Engineering & Technology,* **8**, 8, 190-194.

Nandgude, S. and Tiwari, M., 2023, "Drought prediction: A comprehensive review of different drought prediction models and adopted technologies", *Sustainability,* **15**, 14, 11684. https://doi.org/10.3390/su151411684.

Oyouonalsoud, M. S., Yilmaz, A. G., Abdullah, M., and Abdeljaber, A., 2023, "Drought prediction using artificial intelligence models based on climate data and soil moisture", *Scientific Reports,* **13,** 11839. https://doi.org/10.1038/s41598-023-38849-5.

Ramesh K., and Iyengar R. N., 2017, "A non-Gaussian model for Indian monsoon rainfall", *International Journal of Research in Granthaalayah,* **5**, 4, 88–95. https://doi.org/10.5281/zenodo.556423.

Ramesh K., and Iyengar, R. N., 2017, "Forecasting Indian monsoon rainfall including within-year seasonal variability", International *Journal of Civil Engineering and Technology***,** **8**, 4, 390–399.

Ramesh K., and Iyengar R. N., 2017, "Rainfall distribution and probabilistic forecasting: A review", *Journal of Earth System Science*, **126**, 3, 34. https://doi.org/10.1007/s12040-017-0810-6**.**

Ramesh K., and Kumudha H. R., 2019, "A review on forecasting Indian monsoon rainfall", *International Journal of Innovative Science and Research Technology, Special Issue,* AAM 2019, 9–14.

Smakhtin, V. U. and Hughes, D. A., 2007, "Automated estimation and analyses of meteorological drought characteristics from monthly rainfall data", Environmental *Modelling and Software***, 22**, ***6***, 880-890. https://doi.org/10.1016/j.envsoft.2006.05.013.

Srivastava, A. K., Rajeevan, M., and Kshirsagar, S. R. 2009, "Development of a high-resolution daily gridded rainfall data set for the Indian region", *Current Science*, **96**, 4, 558-562.

Srivastava, A., Ray, K. N., De, U. S. and Mohanty, U. C. 2009, "Defining & predicting agricultural droughts in India using rainfall probability and ENSO index", *Journal of Hydrology*, **370**,1-4,142-155. https://doi.org/10.1016/j.jhydrol.2009. 02.052.

Subrahmanyam, K. V., Cramsenthil, K., 2021, "Prediction of heavy rainfall days over a peninsular Indian station using machine learning algorithms", *Journal of Earth System Science***, 130**, 240, 1-12. https://doi.org/10.1007/s12040-021-01700-2.

Wable, P. S., Jha, M. K., and Shekhar A., 2019, "Comparison of drought indices in a semi-arid river basin of India", *Water Resources Management,* **33**, 1, 75-102. https://doi.org/10.1007/s11269-018-2089-z.

Wilhite, D. A., and Glantz, M. H., 1985, "Understanding the drought phenomenon: The role of definitions", *Water International,* **10**, ***3***, 111–120. https://doi.org/10.1080/02508068508686328.

Wilks, D. S., 2011, "Statistical methods in the atmospheric sciences***"***, *Academic Press,* 3rd ed.

https://www.researchgate.net/publication/356089836_Meteorological_sub-divisions_of_India_and_their_geopolitical_evolution_from_1875_to_2020.